

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:08:41 ; Search time 6729 Seconds  
(without alignments)  
12786.661 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggacagggaggagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gsl:\*  
10: gb\_gsl2:\*  
11: gb\_gsl3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# ALIGNMENTS

RESULT 1  
AK049358  
LOCUS  
DEFINITION

AK049358 1832 bp mRNA linear HTC 03-APR-2004  
Mus musculus ES cells cDNA, RIKEN full-length enriched library,  
clone:C30027E02 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC  
ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert  
sequence.

ACCESSION AK049358  
VERSION AK049358.1 GI:26340087  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
10349636  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
11076861

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

4 The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USPTO)**



天

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:08:41 ; Search time 6729 Seconds  
(without alignments)  
12786.661 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacgaggaggagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_est3: \*  
4: gb\_hic: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_est7: \*  
9: gb\_gsa1: \*  
10: gb\_gsa2: \*  
11: gb\_gsa3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1799.2	97.8	1832	AK049358	AK049358 Mus muscu
2	1785	97.1	1787	AK086940	AK086940 Mus muscu
3	1780	96.8	1896	AK085617	AK085617 Mus muscu
4	1683	91.5	1684	AK031742	AK031742 Mus muscu
5	1528	83.1	1553	AK005532	AK005532 Mus muscu
6	1388.4	75.5	1621	AK082340	AK082340 Mus muscu
7	1352.8	73.6	1376	AK082749	AK082749 Mus muscu
8	1345.6	73.2	1586	AK010702	AK010702 Mus muscu
9	766	41.7	766	CX166520	CX166520 HES2C41
10	761.2	41.4	766	BG974240	BG974240 602843972
11	751.4	40.9	926	BQ922065	BQ922065 AGENCOURT
12	743.8	40.4	793	BI104487	BI104487 602890091
13	735.6	40.0	937	CB203728	CB203728 AGENCOURT
14	733.8	39.9	768	BU522275	BU522275 AGENCOURT
15	733.6	39.9	892	CF553710	CF553710 AGENCOURT
16	728.8	39.6	903	CB320984	CB320984 AGENCOURT
17	722.2	39.3	985	BF785328	BF785328 602111557
18	713	38.8	737	BG873094	BG873094 602794679
19	711.4	38.7	836	BI078951	BI078951 602873203
20	695.2	37.8	970	CF584966	CF584966 AGENCOURT
21	694.2	37.7	699	CX208584	CX208584 MNS15855
22	691.6	37.6	1330	CR626260	CR626260 full-leng

23	691.6	37.6	1338	4	CR625170	CR625170 full-leng
24	691.6	37.6	1384	4	CR603815	CR603815 full-leng
25	691.6	37.6	1393	4	CR605951	CR605951 full-leng
26	691.6	37.6	1461	4	CR606170	CR606170 full-leng
27	691.6	37.6	1657	4	CR620849	CR620849 full-leng
28	691.6	37.6	1659	4	CR593913	CR593913 full-leng
29	691.6	37.6	1667	4	CR605378	CR605378 full-leng
30	686.2	37.3	1316	4	CR614568	CR614568 full-leng
31	682	37.1	696	6	CD353185	CD353185 UI-M-GLO-
32	678.8	36.9	1140	5	EX375141	EX375141 BX375141
33	678.4	36.9	721	10	AY403772	AY403772 Mus muscu
34	677.8	36.9	942	2	BF385997	BF385997 602046176
35	677	36.8	1024	5	BY710119	BY710119 BY710119
36	672	36.5	991	5	BQ876296	BQ876296 AGENCOURT
37	669.4	36.4	886	5	EX383415	EX383415 AGENCOURT
38	663.8	36.1	1136	3	BM545167	BM545167 AGENCOURT
39	659.2	35.8	1694	4	CR608091	CR608091 full-leng
40	652	35.5	665	6	CF732471	CF732471 UI-M-HAO-
41	646.8	35.2	698	2	BI110829	BI110829 602895470
42	641.8	34.9	779	3	BI688404	BI688404 60316338
43	640.6	34.8	944	5	BU510421	BU510421 AGENCOURT
44	638.8	34.7	768	8	DN935262	DN935262 AGENCOURT
45	638.2	34.7	645	2	BB498921	BB498921 BB498921

## ALIGNMENTS

RESULT 1	AK049358	AK049358	1832 bp	mrna	linear	HTC 03-APR-2004
LOCUS	AK049358	Mus musculus ES cells cDNA, RIKEN full-length enriched library,				
DEFINITION	AK049358	ACID: C330027802 product: 1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.				
ACCESSION	AK049358	AK049358				
VERSION	AK049358.1	GI: 26340087				
KEYWORDS	HTC; CAP trapper.					
SOURCE	Mus musculus (house mouse)					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1	Carninci, P. and Hayashizaki, Y.				
AUTHORS	High-efficiency full-length cDNA cloning					
TITLE	Meth. Enzymol. 303, 19-44 (1999)					
JOURNAL	10349636					
PUBMED						
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
AUTHORS	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes					
TITLE	Genome Res. 10 (10), 1617-1630 (2000)					
JOURNAL	11042159					
PUBMED						
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
AUTHORS	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer					
TITLE	Genome Res. 10 (11), 1757-1771 (2000)					
JOURNAL	11076861					
PUBMED						
REFERENCE	4	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.				
AUTHORS	Functional annotation of a full-length mouse cDNA collection					
TITLE	Nature 409, 685-690 (2001)					
JOURNAL						
PUBMED						
REFERENCE	5					

AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	Qy	129	TCCTGCCCGAAGCTAGGCCCCGAGTCCCGCGGGCTTGAGGAGACCGACTGGGTGCGATT	188
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	Ds	149	TCCTGCCCGAAGCTAGGCCCCGAGTCCCGCGGGCTTGAGGAGACCGACTGGGTGCGATT	208
JOURNAL REFERENCE	Nature 420, 563-573 (2002)	Qy	189	GCCAGCAATGCGAAGTGTGCAAGTATGTTCTGTGGAGCTGAAGTCGGCTTTTGAGGA	248
AUTHORS	6 (bases 1 to 1832)	Ds	209	GCCAGCAATGCGAAGTGTGCAAGTATGTTCTGTGGAGCTGAAGTCGGCTTTTGAGGA	268
	Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kohima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saiboh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tgami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Qy	249	AACGGGAAGACCAAGGAAGTGTGACACCGGTATGGCATCTCGACGGGAAGGGCTC	308
	Direct Submission	Ds	269	AACGGGAAGACCAAGGAAGTGTGACACCGGTATGGCATCTCGACGGGAAGGGCTC	328
TITLE	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]	Qy	309	TGAGTCAAGTACACCAAGTCGGAATTCACGGTTAAATGAACTCACTGAGACCAATTTGCAA	368
JOURNAL	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	Ds	329	TGAGTCAAGTACACCAAGTCGGAATTCACGGTTAAATGAACTCACTGAGACCAATTTGCAA	388
	Please visit our web site for further details.	Qy	369	GAGGCTTCTGGAATACAGCTGCAACAGGAGAGACTGGCGAGCAACCGTTTGCARAGGG	428
	URL: http://genome.gsc.riken.jp/	Ds	389	GAGGCTTCTGGAATACAGCTGCAACAGGAGAGACTGGCGAGCAACCGTTTGCARAGGG	448
	URL: http://fantom.gsc.riken.jp/	Qy	429	TATGTCGGAGACCTTTTCAGACGCTGCACAACTAGTCCACAAAGGGGTCAAGTGTGTAT	488
FEATURES	Location/Qualifiers	Ds	449	TATGTCGGAGACCTTTTCAGACGCTGCACAACTAGTCCACAAAGGGGTCAAGTGTGTAT	508
source	1. 1832	Qy	489	GGATATATCCCTATGAGCTGTGGAACGAGACTCTCAGCAGAGGTGGCTCAACGAAGCA	548
	/organism="Mus musculus"	Ds	509	GGATATATCCCTATGAGCTGTGGAACGAGACTCTCAGCAGAGGTGGCTCAACGAAGCA	568
	/mol_type="mRNA"	Qy	549	GTGTGAGCTGCTGTGGAAGATTGGAAGAGTGAATTCAGGACTGGTACAGGACCAACCA	608
	/strain="C57BL/6J"	Ds	569	GTGTGAGCTGCTGTGGAAGATTGGAAGAGTGAATTCAGGACTGGTACAGGACCAACCA	628
	/db_xref="FANTOM_DB:C330027E02"	Qy	609	GGAGGAAGACTGACTCAATTTCTCTGTGCAACACAGCTGCTGAAGGGAAGGACACGAG	668
	/db_xref="taxon:10090"	Ds	629	GGAGGAAGACTGACTCAATTTCTCTGTGCAACACAGCTGCTGAAGGGAAGGACACGAG	688
	/clone="C330027E02"	Qy	669	TTGCTTAGCAGAGCGGTGCTTGGCAAGAGGGGGACATAGCTCTCTGGAGAGGAAGAA	728
	/cell_type="ES cells"	Ds	689	TTGCTTAGCAGAGCGGTGCTTGGCAAGAGGGGGACATAGCTCTCTGGAGAGGAAGAA	748
	/clone_lib="RIKEN full-length enriched mouse cDNA library"	Qy	729	ATCCAAAGAGAGCGCAGCGAGTCAAGGGTCTCCTCAGTGGCAGCAGCAGCAGGAGAA	788
CDS	75..905	Ds	749	ATCCAAAGAGAGCGCAGCGAGTCAAGGGTCTCCTCAGTGGCAGCAGCAGCAGGAGAA	808
	/note="unnamed protein product; 1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE) (SPTR Q9DAU1, evidence: FASTY, 100%ID, 100%length, match=828)	Qy	789	GGAACTGGGGGCTGGGGAGGATGCCAACGCGAGGAGGAGGAGTGTGCAGAAAGGC	848
	putative"	Ds	809	GGAACTGGGGGCTGGGGAGGATGCCAACGCGAGGAGGAGGAGTGTGCAGAAAGGC	868
	/codon_start=1	Qy	849	ATCGCCCTCTCCACACAGCCCCCTGTGATGAGCTGTGAGCCAGCTTAGTGTCTTGAATC	908
	/protein_id="BAC33707.1"	Ds	869	ATCGCCCTCTCCACACAGCCCCCTGTGATGAGCTGTGAGCCAGCTTAGTGTCTTGAATC	928
	/db_xref="GI:26340088"	Qy	909	AAGACCCCTGACTTCAGAGCTTGGGACCGCAGCGCGGCGAGCGCAGCTCCAGCAG	968
	/translation="MESSELAPELCLFLLPFLLLPAPKLGSPAGABETDWR LPSCKEVCYVAVELGKIVGTGKTEIDYGLIDGKSGVKYKVDIPLYELNETSAEVA ICKRLDYSLKHRTGNSRFAGKMSFTFETLHNLVHKGVKVMVDIPLYELNETSAEVA DLKQCDVLVEFEFEVIEDWYRNQBEDLTFELCANHVLKQDTSCLAEBSGKGGDI ASLGKKSKKRSRGVKGSSGSKQRKELGELGEDANAEDEEGVQKASPLPHSPDEL"	Ds	929	AAGACCCCTGACTTCAGAGCTTGGGACCGCAGCAGCGCGCAGCGCAGCTCCAGCAAG	988
	"	Qy	969	GACAGCTGCTGTCCAGCATCAGGTCTCTCCCTTGGCTGTGCCCTTTCCTTCCCTTCAA	1028
ORIGIN		Ds	989	GACAGCTGCTGTCCAGCATCAGGTCTCTCTCCCTTGGCTGTGCCCTTTCCTTCCCTTCAA	1048
	Query Match 97.8%; Score 1799.2; DB 4; Length 1832;	Qy	1029	CAACAGCAAGAGGTGGAAGGATCTGGGGTGTCTGGGAGACGGCACCCCAAGGGGAAGGA	1088
	Best Local Similarity 99.8%; Pred. No. 0;	Ds	1049	CAACAGCAAGAGGTGGAAGGATCTGGGGTGTCTGGGAGACGGCACCCCAAGGGGAAGGA	1108
	Matches 1801; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	Qy	1089	GGAGGAGAGAGGAGCTCTTTTCTACAGTCCCTCCCTCAGAGCTCCGGGGTCCACC	1148
		Ds	1109	GGAGGAGAGAGGAGCTCTTTTCTACAGTCCCTCCCTCAGAGCTCCGGGGTCCACC	1168
		Qy	1149	CAGCATCCCCAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGCATGAGGCAAT	1208
		Ds	1169	CAGCATCCCCAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGCATGAGGCAAT	1228
		Qy	1209	AAGATGCTCACAGCGCCCCCTTTCAGCCAGGAAGGACTCCGCTGCGAGCCAGG	1268
		Ds	1269	AAGATGCTCACAGCGCCCCCTTTCAGCCAGGAAGGACTCCGCTGCGAGCCAGG	1288



DLKQCDVLVEEFEEVIEDWYRNHQBEDLTFELCANHLVKGKDTSLAERWSGKGGDI  
ASLGGKSKKRSVGVKSSSGSKQKRELGGLEDANAEBEEGVQKASPLPHSPDEL

ORIGIN

Query Match		97.1%;	Score 1785;	DB 4;	Length 1787;		
Best Local Similarity		100.0%;	Pred. No. 0;				
Matches 1785;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
Qy	35	GCTCTGGGTCGGGTCATGGAGTCCATGTCTGAGCTCGCGCCCCGCTGCTCTTAT	94				
Db	2	GCTCTGGGTCGGGTCATGGAGTCCATGTCTGAGCTCGCGCCCCGCTGCTCTTAT	61				
Qy	95	TTCTTTTGTGCTGCTCTTCGCTGCTGCTCTTCCTGCTGCGGAGCTAGGCCGAGTC	154				
Db	62	TTCTTTTGTGCTGCTCTTCGCTGCTGCTCTTCCTGCTGCGGAGCTAGGCCGAGTC	121				
Qy	155	CCGCGGGGCTGAGGACACGACTGGGTGCGATTGCCCCAGCAATGCGAAGTGTGCAAGT	214				
Db	122	CCGCGGGGCTGAGGACACGACTGGGTGCGATTGCCCCAGCAATGCGAAGTGTGCAAGT	181				
Qy	215	ATGTTGCTGTGGAGCTGAAGTCGGCTTTTGAGGAAACGGGAAAGACCAAGGAAGTGATTG	274				
Db	182	ATGTTGCTGTGGAGCTGAAGTCGGCTTTTGAGGAAACGGGAAAGACCAAGGAAGTGATTG	241				
Qy	275	ACACCGGCTATGGCATCTCTGGAACGGGAGGCTCTGGAGTCMAGTACACCAAGTCGGACT	334				
Db	242	ACACCGGCTATGGCATCTCTGGAACGGGAGGCTCTGGAGTCMAGTACACCAAGTCGGACT	301				
Qy	335	TACGGTTAAATTGAAGTCACTGAGACCAATTTGCAAGAGCTTCTGGACTACAGCCTGCACA	394				
Db	302	TACGGTTAAATTGAAGTCACTGAGACCAATTTGCAAGAGCTTCTGGACTACAGCCTGCACA	361				
Qy	395	AGGAGGAGCTGCGCAGCAACCGGTTTCCAAAGGGTATGTCCGAGACCTTTTGAGACGCTGC	454				
Db	362	AGGAGGAGCTGCGCAGCAACCGGTTTCCAAAGGGTATGTCCGAGACCTTTTGAGACGCTGC	421				
Qy	455	ACAACCTAGTCCACAAAGGGGTCAAGGTGTGATGATATCCCTATGAGCTGTGGAAAG	514				
Db	422	ACAACCTAGTCCACAAAGGGGTCAAGGTGTGATGATATCCCTATGAGCTGTGGAAAG	481				
Qy	515	AGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGACGTCTGGTGAAGAGTTTG	574				
Db	482	AGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGACGTCTGGTGAAGAGTTTG	541				
Qy	575	AAGAGGTGATTGAGGACTGTGTACAGGAAACCAAGGAGGAGAGCTGACTGAATTCCTCT	634				
Db	542	AAGAGGTGATTGAGGACTGTGTACAGGAAACCAAGGAGGAGAGCTGACTGAATTCCTCT	601				
Qy	635	GTGCCAACCAAGCTCTGAGGAGGAGGACAGAGTTCGCTAGCAGAGGGTGTCTGGCA	694				
Db	602	GTGCCAACCAAGCTCTGAGGAGGAGGACAGAGTTCGCTAGCAGAGGGTGTCTGGCA	661				
Qy	695	AGAGGGGGACATAGCTCTCCCTGGGAGGAGAAATCCAGAGAGCGACGCGAGTCA	754				
Db	662	AGAGGGGGACATAGCTCTCCCTGGGAGGAGAAATCCAGAGAGCGACGCGAGTCA	721				
Qy	755	AGGGCTCCTCAGTGGCAGCAGCAAGCAGAGAAAGGAACTGGGGGGCTGGGGAGGATG	814				
Db	722	AGGGCTCCTCAGTGGCAGCAGCAAGCAGAGAAAGGAACTGGGGGGCTGGGGAGGATG	781				
Qy	815	CCACGCGGAGGAGGAGGAGGTGTGCAAGGAGTGCCTCCACACAGCCCCCTG	874				
Db	782	CCACGCGGAGGAGGAGGAGGTGTGCAAGGAGTGCCTCCACACAGCCCCCTG	841				
Qy	875	ATGAGCTGTGAGCCAGCTTGTCTTGAATCAAGACCCCTGACTTCAGAGCTTGGA	934				
Db	842	ATGAGCTGTGAGCCAGCTTGTCTTGAATCAAGACCCCTGACTTCAGAGCTTGGA	901				
Qy	935	CACGACACGCGCAGCGCAGCTCCAGCAAGGACAGCTGTGTCCAGCATCAGGTCT	994				
Db	902	CACGACACGCGCAGCGCAGCTCCAGCAAGGACAGCTGTGTCCAGCATCAGGTCT	961				

Qy	995	CCTCCCTTGGCTGTGCCCCCTTCTCTCCCTTGAACACAGCAAGAGTGAAGGATCTGG	1054				
Db	962	CCTCCCTTGGCTGTGCCCCCTTCTCTCCCTTGAACACAGCAAGAGTGAAGGATCTGG	1021				
Qy	1055	GGTGTGGGAGACGGGACCCCAAGGGAAGAGGAGGAGGAGAGGAGGAGGAGGAGGAGG	1114				
Db	1022	GGTGTGGGAGACGGGACCCCAAGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1081				
Qy	1115	TACACAGTCCCTCTCAGGAGCTCGGGGTCCACCCAGCATCCCCAGGCTCGAGATCCAGGC	1174				
Db	1082	TACACAGTCCCTCTCAGGAGCTCGGGGTCCACCCAGCATCCCCAGGCTCGAGATCCAGGC	1141				
Qy	1175	TCCTGACATGAAGAGCTGAAGAGCATGAGGCACATAAGATGCTCACACGCGCCCCCTTCAG	1234				
Db	1142	TCCTGACATGAAGAGCTGAAGAGCATGAGGCACATAAGATGCTCACACGCGCCCCCTTCAG	1201				
Qy	1235	CCAGGAAGGACTCCGTGACGCTCAGCAGCAGGCGCTGCTCTTCTTCCACCAAGCATTT	1294				
Db	1202	CCAGGAAGGACTCCGTGACGCTCAGCAGCAGGCGCTGCTCTTCTTCCACCAAGCATTT	1261				
Qy	1295	CTCTTCTGCTGTGCTTGTGCGATGTTAAATTCGAGAACTTCCAGGACAAACTCGGGTGT	1354				
Db	1262	CTCTTCTGCTGTGCTTGTGCGATGTTAAATTCGAGAACTTCCAGGACAAACTCGGGTGT	1321				
Qy	1355	GGCACAAGGGGCTGGACGCGCAGAGCCAGACGACGCGCAGAGACTGCAGAGAGGCGACCT	1414				
Db	1322	GGCACAAGGGGCTGGACGCGCAGAGCCAGACGCGCAGAGACTGCAGAGAGGCGACCT	1381				
Qy	1415	GACCTAAACCCCTTGGAAAGCCAAATCTGCAAGTTCCTGTCACCCATCTCTCTGAGGA	1474				
Db	1382	GACCTAAACCCCTTGGAAAGCCAAATCTGCAAGTTCCTGTCACCCATCTCTCTGAGGA	1441				
Qy	1475	CGCTCATGCTCTGCCCGCCGCTTCTCCAGGGGTACAGAGTAACACCTTTTGGGCTTT	1534				
Db	1442	CGCTCATGCTCTGCCCGCCGCTTCTCCAGGGGTACAGAGTAACACCTTTTGGGCTTT	1501				
Qy	1535	TCGGTTGGTTCCTGGGTCCTCATCAGCCTCCAGAGTGTCCCTCATCGATCTTTTTCG	1594				
Db	1502	TCGGTTGGTTCCTGGGTCCTCATCAGCCTCCAGAGTGTCCCTCATCGATCTTTTTCG	1561				
Qy	1595	CTTTGTCCCAATCCAGGGGCTGGAAGGCGATCACCATCATTTGGAGGCTTAACTGTCT	1654				
Db	1562	CTTTGTCCCAATCCAGGGGCTGGAAGGCGATCACCATCATTTGGAGGCTTAACTGTCT	1621				
Qy	1655	AGTTACTAGGAGTGTCTGGAGCGCCCGGGGTGGTTGGGTAATCACTACCTGGCTCT	1714				
Db	1622	AGTTACTAGGAGTGTCTGGAGCGCCCGGGGTGGTTGGGTAATCACTACCTGGCTCT	1681				
Qy	1715	CAGCCTTAAACACTGACGCCCCCTTAAATACAGTTTCTTCTGTGTGTGACTCCACGCG	1774				
Db	1682	CAGCCTTAAACACTGACGCCCCCTTAAATACAGTTTCTTCTGTGTGTGACTCCACGCG	1741				
Qy	1775	CCACACACACACATATAATTTTCGATGCTGTTTCATAACTG	1819				
Db	1742	CCACACACACACATATAATTTTCGATGCTGTTTCATAACTG	1786				

RESULT 3

AK085617

LOCUS

DEFINITION

AK085617

VERSION

KEYWORDS

SOURCE

ORGANISM

AK085617

AK085617.1

GI:26351704

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1

AK085617 1896 bp mRNA linear HTC 03-APR-2004  
Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched  
library, clone:D630047M13 product:1600025D17RIK PROTEIN (PUTATIVE  
RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full  
insert sequence.

AK085617

AK085617.1

GI:26351704

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1







Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.sgc.riken.jp/>  
URL: <http://fantom.sgc.riken.jp/>  
Location/Qualifiers  
1. 1684

# FEATURES

## source

/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:6230400B22"  
/db\_xref="taxon:10090"  
/clone="6230400B22"  
/tissue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="11 days embryo"  
<1. 757

## CDS

/note="unnamed protein product; 1600025D17RIK PROTEIN  
(PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA  
1600025D17 GENE) (SPTR|Q9DAU1, evidence: FASTY, 100%ID,  
100%length, match=828)  
putative"  
/codon\_start=2  
/protein\_id="BAC27534.1"  
/db\_xref="GI:26327581"  
/translation="PAKLGPSAGABETDWRLPKSCVCKVAVELKSAFETGKT  
KEVDTGYGLDGGSGVMDYKPSDLRLIEVETICKLLDYSLKHETGSGNRFKNGS  
ETFEITLNLVHKGVKVDYPIYELWNETSABADLKQKQDVLVEEFREVIDWYRNHQ  
EEDLTFELCANHLVKGKDTSLAERWSGKGDIASLGGKKSKKRSVGVKSGSSGSKQ  
RKELGGLGEDANAEEREGVQKASPLPHSPPEL"

## ORIGIN

Query Match 91.5%; Score 1683; DB 4; Length 1684;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 130 CCTGCCCGGAGCTAGGCCGAGTCCCGCGGGCTGAGGACCGGACTGGGTCGGATTG 189  
Db 2 CCTGCCCGGAGCTAGGCCGAGTCCCGCGGGCTGAGGACCGGACTGGGTCGGATTG 61  
Qy 190 CCCAGCAAAATCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTGGCTTTTGAGAA 249  
Db 62 CCCAGCAAAATCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTGGCTTTTGAGAA 121  
Qy 250 ACGGAAAGCAAGGAAGTATTGACACCGGCTATGGCATCTGGACGGGAAGGCTCT 309  
Db 122 ACGGAAAGCAAGGAAGTATTGACACCGGCTATGGCATCTGGACGGGAAGGCTCT 181  
Qy 310 GGAGTCAAGTACCAAGTCTGAGCTTACGGTTAATTGAAGTCACTGAGACCATTTGCAAG 369  
Db 182 GGAGTCAAGTACCAAGTCTGAGCTTACGGTTAATTGAAGTCACTGAGACCATTTGCAAG 241  
Qy 370 AGGCTTCTGGAAGTACAGCTGTCACAAAGGAGGAGTCTGGCAGCAACCGGTTTGCAGGGT 429  
Db 242 AGGCTTCTGGAAGTACAGCTGTCACAAAGGAGGAGTCTGGCAGCAACCGGTTTGCAGGGT 301  
Qy 430 ATGTGCGAGACCTTTGAGACGCTGCAACCTAGTCCACAAAGGGGTCAAGGTGCTATG 489  
Db 302 ATGTGCGAGACCTTTGAGACGCTGCAACCTAGTCCACAAAGGGGTCAAGGTGCTATG 361  
Qy 490 GATATCCCTATGAGCTGTGGAACAGACCTCAGCAGAGGTGGTGGCTCAAGAACGAG 549  
Db 362 GATATCCCTATGAGCTGTGGAACAGACCTCAGCAGAGGTGGTGGCTCAAGAACGAG 421  
Qy 550 TGTGACGTGCTGTGGAAGAGTTTCAAGAGAGTGTGTTGAGGACTGGTACAGGAACCAACG 609  
Db 422 TGTGACGTGCTGTGGAAGAGTTTCAAGAGAGTGTGTTGAGGACTGGTACAGGAACCAACG 481  
Qy 610 GAGGAAGACCTGACTGAATTCCTCTGTGCCAACACAGTGTGGAAGGGAAGGACACAGT 669  
Db 482 GAGGAAGACCTGACTGAATTCCTCTGTGCCAACACAGTGTGGAAGGGAAGGACACAGT 541

Qy 670 TGCCTACGAGCGGTGGTCTTGGCAAGAGGGGGACATAGCCTCTCTCGGAGGGAGAAA 729  
Db 542 TGCCTACGAGCGGTGGTCTTGGCAAGAGGGGGACATAGCCTCTCTCGGAGGGAGAAA 601  
Qy 730 TCCAAGAAGAAGCGCAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAGCAGAGGAAG 789  
Db 602 TCCAAGAAGAAGCGCAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAGCAGAGGAAG 661  
Qy 790 GAATCGGGGGCTTGGGGAGGATGCCAAGCCGAGGAGGAGGGGTGTCAGAAAGCA 849  
Db 662 GAATCGGGGGCTTGGGGAGGATGCCAAGCCGAGGAGGAGGGGTGTCAGAAAGCA 721  
Qy 850 TCGCCCTTCCACACAGCCCTTGTAGTGTGAGCCAGCTTAGTGTCTTGAATCA 909  
Db 722 TCGCCCTTCCACACAGCCCTTGTAGTGTGAGCCAGCTTAGTGTCTTGAATCA 781  
Qy 910 AGACCTCTGATTCAGAGCTTGGGACACGACGCGAGCGCAGCGAGCTCCAGCAAGG 969  
Db 782 AGACCTCTGATTCAGAGCTTGGGACACGACGCGAGCGCAGCGAGCTCCAGCAAGG 841  
Qy 970 ACAGCTCTGTCCAGCATCAGGTCTCTCTCCCTTGGCTGTGCCCTTTCTTCCCTTGAAC 1029  
Db 842 ACAGCTCTGTCCAGCATCAGGTCTCTCTCCCTTGGCTGTGCCCTTTCTTCCCTTGAAC 901  
Qy 1030 AACAGCAAGAGGTGGAAGGATCTGGGTGTCTGGGAGACGGCACCCCAAGGGAAGGAG 1089  
Db 902 AACAGCAAGAGGTGGAAGGATCTGGGTGTCTGGGAGACGGCACCCCAAGGGAAGGAG 961  
Qy 1090 GAGGAGCAGAAGGAGCTCTCTTCTACAGTCCCTCTACAGCTCCCGGCTCCGCGTCCACCC 1149  
Db 962 GAGGAGCAGAAGGAGCTCTCTTCTACAGTCCCTCTACAGCTCCCGGCTCCGCGTCCACCC 1021  
Qy 1150 AGCATCCCGAGGTGAGATCCAGGCTCTGTGACATGGAAGCTGAAGAGCATGAGGCACATA 1209  
Db 1022 AGCATCCCGAGGTGAGATCCAGGCTCTGTGACATGGAAGCTGAAGAGCATGAGGCACATA 1081  
Qy 1210 AGATGCTCACAGCGCCCTTTCAGCAGGAAGGACTCCGTGTCAGCTTCAGAGCCAGGC 1269  
Db 1082 AGATGCTCACAGCGCCCTTTCAGCAGGAAGGACTCCGTGTCAGCTTCAGAGCCAGGC 1141  
Qy 1270 CTGCTCTCTCTTCCACCAAGCATTTCTCTGCTGTGCTTCTGCGATGTAATTCGA 1329  
Db 1142 CTGCTCTCTCTTCCACCAAGCATTTCTCTGCTGTGCTTCTGCGATGTAATTCGA 1201  
Qy 1330 GAATCTTCAGGACAAAATCTCGGGTGTGGCAAAAAGGGGTGGACGCGAGAGCCAGCCAC 1389  
Db 1202 GAATCTTCAGGACAAAATCTCGGGTGTGGCAAAAAGGGGTGGACGCGAGAGCCAGCCAC 1261  
Qy 1390 GCCAGAGACTGCAGAGGGGCACTGACCTTAACCCCTCTGGAAGGCCAATCTGCAAGTTC 1449  
Db 1262 GCCAGAGACTGCAGAGGGGCACTGACCTTAACCCCTCTGGAAGGCCAATCTGCAAGTTC 1321  
Qy 1450 CGTGTCCACCACTCTCTGAGGACGCTCATGCTCTGCCAGCCCTTCTCCAGGGCT 1509  
Db 1322 CGTGTCCACCACTCTCTGAGGACGCTCATGCTCTGCCAGCCCTTCTCCAGGGCT 1381  
Qy 1510 ACCAGAGTAAACACTTTTGGCTTTTGGTTTGGTTTCTTGGGTCTCATCAGCTTCCAGA 1569  
Db 1382 ACCAGAGTAAACACTTTTGGCTTTTGGTTTGGTTTCTTGGGTCTCATCAGCTTCCAGA 1441  
Qy 1570 GTGTCCCTCATCGATCTTTTTCCTTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 1629  
Db 1442 GTGTCCCTCATCGATCTTTTTCCTTTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC 1501  
Qy 1630 ACCATCATTTGGAGCTTAACTGTGCTACTTAGAGGTGCTGGAGCGCCCGGGTTGG 1689  
Db 1502 ACCATCATTTGGAGCTTAACTGTGCTACTTAGAGGTGCTGGAGCGCCCGGGTTGG 1561  
Qy 1690 TTTGGGTAACTCACTCACTGCTCTGAGCTTCTAAACACTCAGCCCTTAAATACAGTTC 1749  
Db 1562 TTTGGGTAACTCACTCACTGCTCTGAGCTTCTAACTCAGCCCTTAACTACAGTTC 1621



Qy 309 TGGAGTCAAGTACACCAAGTCGAGTCTTACGGTTAAATTGAAGTCACTGAGACCAATTGCAA 368  
 Db 326 TGGAGTCAAGTACACCAAGTCGAGTCTTACGGTTAAATTGAAGTCACTGAGACCAATTGCAA 385  
 Qy 369 GAGGCTTCTGAGTACAGCCTGCAACAGGAGAGGACTGCGCAGCAACCGGTTTGCCAGGG 428  
 Db 386 GAGGCTTCTGAGTACAGCCTGCAACAGGAGAGGACTGCGCAGCAACCGGTTTGCCAGGG 445  
 Qy 429 TATGTCGAGACCTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGGTGAAT 488  
 Db 446 TATGTCGAGACCTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGGTGAAT 505  
 Qy 489 GGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGAGTGCGTGAACCTCAAGAGCA 548  
 Db 506 GGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGAGTGCGTGAACCTCAAGAGCA 565  
 Qy 549 GTGTGAGCTGTGTTGGAAGAGTTTGAAGAGTGATTTGAGGACGTGTTACAGGACCTACAGCAACCA 608  
 Db 566 GTGTGAGCTGTGTTGGAAGAGTTTGAAGAGTGATTTGAGGACGTGTTACAGGACCTACAGCAACCA 625  
 Qy 609 GGAGGAAGACCTGACTGAATTTCTCTGTGCAACACGCTGCTGAAGGGAAGGACACGAG 668  
 Db 626 GGAGGAAGACCTGACTGAATTTCTCTGTGCAACACGCTGCTGAAGGGAAGGACACGAG 685  
 Qy 669 TTGCTAGCAGAGCGGTGCTGCGCAAGAGGGGACATAGCCTCCTCGGAGGGAAGAA 728  
 Db 686 TTGCTAGCAGAGCGGTGCTGCGCAAGAGGGGACATAGCCTCCTCGGAGGGAAGAA 745  
 Qy 729 ATCCAAGAGAGCGCAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAGCAGAGGAA 788  
 Db 746 ATCCAAGAGAGCGCAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAGCAGAGGAA 805  
 Qy 789 GGAATCGGGGCTCGGGGAGGATGCAACCGCAGGAGGAGGAGGTGTGCAAGAGGC 848  
 Db 806 GGAATCGGGGCTCGGGGAGGATGCAACCGCAGGAGGAGGAGGTGTGCAAGAGGC 865  
 Qy 849 ATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTGTAGTTCCTTGAATC 908  
 Db 866 ATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTGTAGTTCCTTGAATC 925  
 Qy 909 AAGACCCCTGACTTCAGAGCTTGGGACGACGAGCGGAGCGGAGCGGAGTCCAGCAAG 968  
 Db 926 AAGACCCCTGACTTCAGAGCTTGGGACGACGAGCGGAGCGGAGCGGAGTCCAGCAAG 985  
 Qy 969 GACAGCTGCTGTCCAGCATCAGGTCTCTCTCTTGGCTGTGCCCCCTTCTTCCCTTGAA 1028  
 Db 986 GACAGCTGCTGTCCAGCATCAGGTCTCTCTCTTGGCTGTGCCCCCTTCTTCCCTTGAA 1045  
 Qy 1029 CAACAGCAAGAGTGGAGGATCTGGGGTGTGGGAGACGCGACCCCAAGGGGAAGGGA 1088  
 Db 1046 CAACAGCAAGAGTGGAGGATCTGGGGTGTGGGAGACGCGACCCCAAGGGGAAGGGA 1105  
 Qy 1089 GGAGGAGCAGAGCAGCTCTCTTCTACAGTCCCGCTCAGCAGCTCCGGGTCCACC 1148  
 Db 1106 GGAGGAGCAGAGCAGCTCTCTTCTACAGTCCCGCTCAGCAGCTCCGGGTCCACC 1165  
 Qy 1149 CAGCATCCCAAGGCTGAGATCCAGCTCTGACATGGAAGCTGAAGAGCATGAGGACAT 1208  
 Db 1166 CAGCATCCCAAGGCTGAGATCCAGCTCTGACATGGAAGCTGAAGAGCATGAGGACAT 1225  
 Qy 1209 AAGATGCTCAGCAGCGCCCCCTTCCAGCAGGAAGGACTCCGTGCGAGCTCAGCAGCCAGG 1268  
 Db 1226 AAGATGCTCAGCAGCGCCCCCTTCCAGCAGGAAGGACTCCGTGCGAGCTCAGCAGCCAGG 1285  
 Qy 1269 CTTGCTCTTCTTCCACCAAGCATCTCTTCTCTGCTGCTCTTCTGAGTGGTAAATTCG 1328  
 Db 1286 CTTGCTCTTCTTCCACCAAGCATCTCTTCTCTGCTGCTCTTCTGAGTGGTAAATTCG 1345  
 Qy 1329 AGAATCTCCAGGACAACTCGGGTGTGGCACAAGGGGCTGGAGCCAGAGCCAGAGCCA 1388  
 Db 1346 AGAATCTCCAGGACAACTCGGGTGTGGCACAAGGGGCTGGAGCCAGAGCCAGAGCCA 1405

Qy 1389 CGCCAGAGACTGCAGAGAGGCGACCTGACCTAACCCCTCGAAGGCCAATCTGCAGTTC 1448  
 Db 1406 CGCCAGAGACTGCAGAGAGGCGACCTGACCTAACCCCTCGAAGGCCAATCTGCAGTTC 1465  
 Qy 1449 CCGTGTCCACCCACTCTCTCTGAGGAGCGCTCATGCTCTGCCAGCCCTTCTCCAGGGC 1508  
 Db 1466 CCGTGTCCACCCACTCTCTCTGAGGAGCGCTCATGCTCTGCCAGCCCTTCTCCAGGGC 1525  
 Qy 1509 TACCAGAGTAAACACCTTTTGGCCTTTC 1536  
 Db 1526 TACCAGAGTAAACACCTTTTGGCCTTTC 1553

RESULT 6  
 AK082340  
 LOCUS  
 DEFINITION  
 Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length enriched library, clone:C230039G09 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.  
 ACCESSION  
 VERSION  
 AK082340.1 GI:26349662  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 10349636  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 11042159  
 3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 11076861  
 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 1621)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nomazaki, R., Ohno, M., Oheato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/.

# FEATURES

Source  
1. .1621  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:C230039G09"  
/db\_xref="taxon:10090"  
/clone="C230039G09"  
/tissue\_type="cerebellum"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="0 day neonate"  
230. .685  
/note="unnamed protein product; 1600025D17RIK PROTEIN  
(PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA  
1600025D17 GENE) (SPTR|Q9DAU1, evidence: FASTY, 100%ID,  
100%length, match=828)  
putative"  
/codon\_start=1  
/protein\_id="BAC38471.1"  
/db\_xref="GI:26349663"  
/translation="MSFTFETLNLVHKVYVMDIPELWNETSAEVDLKKQCDVL  
VEEFVEIOWRHHQBEDLTFELCANHLKQDTSCLAEKRWGSKGDIASLGKSKK  
KRSRGVSGSSGSKQKRELGLGEDANAEIEGVQKASPLPHSPDDEL"

## ORIGIN

Query Match 75.5%; Score 1388.4; DB 4; Length 1621;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1400; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
420 TGCCAAAGGGTATGTCGGAGACCTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAA 479  
220 TGCGAAGGGTATGTCGGAGACCTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAA 279  
480 GGTGGTGATGATATCCCTTATAGCTGTGGAAAGAGACCTTCAGACAGAGGTGGCTGACCT 539  
280 GGTGGTGATGATATCCCTTATAGCTGTGGAAAGAGACCTTCAGACAGAGGTGGCTGACCT 339  
540 CAAGAAGCAGTGTGACGTCGTGGTGGAGAGTTTGAAGAGGTGATTGAGGACTGGTACAG 599  
340 CAAGAAGCAGTGTGACGTCGTGGTGGAGAGTTTGAAGAGGTGATTGAGGACTGGTACAG 399  
600 GAACACACAGAGGAAGACCTGACGTAATTCCTCTGTGCCAACCAACGCTGCTGAAGGGA 659  
400 GAACACACAGAGGAAGACCTGACGTAATTCCTCTGTGCCAACCAACGCTGCTGAAGGGA 459  
660 GGAACAGAGTTGCTTAGCAGACGGTGTCTGGCAAGAGGGGACATAGACCTCCCTGGG 719  
460 GGAACAGAGTTGCTTAGCAGACGGTGTCTGGCAAGAGGGGACATAGACCTCCCTGGG 519  
720 AGGAAGAAATCCAAAGAGAGCCGAGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAA 779  
520 AGGAAGAAATCCAAAGAGAGCCGAGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAA 579  
780 GCAGAGGAAGAACTGGGGGGCTTGGGGAGGATGCCAAACGCCGAGGAGGAGGGTGT 839

580	GCAGAGGAAGAACTGGGGGGCTTGGGGAGGATGCCAAACGCCGAGGAGGAGGGTGT	639	
840	GCAGAAGGCATCGCCCTCCACACACAGCCCTCGATGAGCTGTGAGCCACGCTTAGTGT	899	
640	GCAGAAGGCATCGCCCTCCACACACAGCCCTCGATGAGCTGTGAGCCACGCTTAGTGT	699	
900	CTTTGAATCAAGACCCCTGACTTTTCAGAGCTTTGGGACACGACGCGAGCGACGCGAGC	958	
700	CTTTGAATCAAGACCCCTGACTTTTCAGAGCTTTGGGACACGACGCGAGCGACGCGAGC	759	
959	CTCAGCAAGACGAGCTGTCTCAGCATCAGGTCTCTCCCTTGGCTGTGCTCTTCC	1018	
760	CTCAGCAAGACGAGCTGTCTCAGCATCAGGTCTCTCCCTTGGCTGTGCTCTTCC	819	
1019	TTCCCTTTGAACAACAGCAAGAGGTGGAGGATCTGGGGTGTCTGGGACACGCAACCCCAA	1078	
820	TTCCCTTTGAACAACAGCAAGAGGTGGAGGATCTGGGGTGTCTGGGACACGCAACCCCAA	879	
1079	GGGAAGAGGAGGAGGAGCAGAGGAGCTCTCTTTCTACACAGTCCCTCCCTCACGAGCTCC	1138	
880	GGGAAGAGGAGGAGGAGCAGAGGAGCTCTCTTTCTACACAGTCCCTCCCTCACGAGCTCC	939	
1139	GGGTTCACCCAGCATCCCAAGCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCA	1198	
940	GGGTTCACCCAGCATCCCAAGCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCA	999	
1199	TGAGGCACATAGATGCTCACAGCGCCCTTCAGCAGCAAGAGGAGCTCCGTGAGCGCTC	1258	
1000	TGAGGCACATAGATGCTCACAGCGCCCTTCAGCAGCAAGAGGAGCTCCGTGAGCGCTC	1059	
1259	AGCAGCCAGCGCTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGGTCTCTTCTCGGAT	1318	
1060	AGCAGCCAGCGCTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGGTCTCTTCTCGGAT	1119	
1319	GGTAAATTCAGAACTCTCCAGGACAAACTCGGGTGTGGCAAAAGGGGCTGGAGCGCAGA	1378	
1120	GGTAAATTCAGAACTCTCCAGGACAAACTCGGGTGTGGCAAAAGGGGCTGGAGCGCAGA	1179	
1379	GCCAGCCAGCGCAGAGAGCTGCAGAGAGGAGCCTGACCTTAAACCCCTTGGAAGGCCAA	1438	
1180	GCCAGCCAGCGCAGAGAGCTGCAGAGAGGAGCCTGACCTTAAACCCCTTGGAAGGCCAA	1239	
1439	CTGCACTTCCCGTGTCCACCCACTCTCTCTGAGAGCGCTCATGCTCTGCCAGCGCTT	1498	
1240	CTGCACTTCCCGTGTCCACCCACTCTCTCTGAGAGCGCTCATGCTCTGCCAGCGCTT	1299	
1499	CTCCAGGGGTACAGAGTAAACACCTTTTGGCCCTTTCGGTTTGGTCTCTGGGTCTCAT	1558	
1300	CTCCAGGGGTACAGAGTAAACACCTTTTGGCCCTTTCGGTTTGGTCTCTGGGTCTCAT	1359	
1559	CAGCTCCAGAGTGTCCCTCATCGATCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTT	1618	
1360	CAGCTCCAGAGTGTCCCTCATCGATCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTT	1419	
1619	GGAGGCGCATCACCATCAATTGGAGGCTTAACTGTCACTAGTACTAGGAGGTCTGGAGCG	1678	
1420	GGAGGCGCATCACCATCAATTGGAGGCTTAACTGTCACTAGTACTAGGAGGTCTGGAGCG	1479	
1679	CCCGGGTGGTGGGGTAACTCACTCACTGGCTCTCAGCCCTTCTAATGAGAGCCCTT	1738	
1480	CCCGGGTGGTGGGGTAACTCACTCACTGGCTCTCAGCCCTTCTAATGAGAGCCCTT	1539	
1739	TAATACAGTTCCTCTCTGTGGTGTACTCCACGCCCCCAACACACACACACATAAATAT	1798	
1540	TAATACAGTTCCTCTCTGTGGTGTACTCCACGCCCCCAACACACACACACATAAATAT	1599	
1799	TTGATGCTGTTTCAATGCT 1820		
1600	TTGATGCTGTTTCAATGCT 1621		

AK082749 1376 bp mRNA linear HTC 03-APR-2004  
 LOCUS Mus musculus ES cells cDNA, Riken full-length enriched library,  
 DEFINITION clone:C330002N05 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC  
 ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert  
 sequence.  
 AK082749  
 VERSION AK082749.1 GI:26349918  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, A.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watahiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multiplexed sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 11076861  
 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 1376)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tegami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,  
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/.

FEATURES  
 Location/Qualifiers  
 1..1376  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="PANTOM.DB:C330002N05"  
 /db\_xref="taxon:10090"  
 /clone="C330002N05"  
 /cell\_type="ES cells"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 <1..521  
 /note="unnamed protein product; 1600025D17RIK PROTEIN  
 (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA  
 1600025D17 GENE) (SPTR[Q9DAU1, evidence: FASTY, 100%ID,  
 100%length, match=828)  
 putative"  
 /codon\_start=3  
 /protein\_id="BAC38599.1"  
 /db\_xref="GI:26349919"  
 /translation="KRLLDYSLHKERTGSNRFPAKCMSETFETLHNLVHKGVKVMWDIP  
 YELWNETSAEVADLLKQCDVLVEFEFVEDWYRNHQBEDLTFELCAHNLVKGKDTSC  
 LAERWSGKGDIAISLGGKKKSKGRSGVSKSGSSGSKQRKELGGLEDANAEABEEGVQK  
 ASPLPHSPDDEL"

CDS  
 365 GCAAGAGGCTTCTGGACTACAGCCTGCAACAGGAGGAGGACTGGCAGCAACCGGTTTGCCA 424  
 1 GCAAGAGGCTTCTGGACTACAGCCTGCAACAGGAGGAGGACTGGCAGCAACCGGTTTGCCA 60  
 425 AGGGTATGTCGGAGACCTTTGAGAGCGTGCACAACTAGTCCCAAAAGGGGTCAAGGTGG 484  
 61 AGGGTATGTCGGAGACCTTTGAGAGCGTGCACAACTAGTCCCAAAAGGGGTCAAGGTGG 120  
 485 TGATGGATATCCCTATGAGCTGTGGAAAGAGAGACTCAGCAGAGGTGCTGACCTCAAGA 544  
 121 TGATGGATATCCCTATGAGCTGTGGAAAGAGAGACTCAGCAGAGGTGCTGACCTCAAGA 180  
 545 AGCAGTGTGAGCTGCTGCTGGAAGAGGTTTGAAGAGGTTGATTTGAGGACTGGTACAGGAAC 604  
 181 AGCAGTGTGAGCTGCTGCTGGAAGAGGTTTGAAGAGGTTGATTTGAGGACTGGTACAGGAAC 240  
 605 ACCAGGAGGAAGACCTGACTGAATTTCTCTGTGCAACACCTGCTGCAAGGGAAGGACA 664  
 241 ACCAGGAGGAAGACCTGACTGAATTTCTCTGTGCAACACCTGCTGCAAGGGAAGGACA 300  
 665 CGAGTTGCTAGCAGAGCGGTGCTGTGCAAGAGGGAATAGCTCTCCCTGGGAGGGA 724  
 301 CGAGTTGCTAGCAGAGCGGTGCTGTGCAAGAGGGAATAGCTCTCCCTGGGAGGGA 360  
 725 AGAATCAAGAAGAGCGGAGGTCAGAGGCTCTCCAGTGGCAGCAGCAGCAGCAG 784  
 361 AGAATCAAGAAGAGCGGAGGTCAGAGGCTCTCCAGTGGCAGCAGCAGCAGCAG 420  
 785 GGAAGGAACCTGGGGGGCTGGGGGAGGATGCAACCGCGGAGGAGGAGGAGGTGTGCAGA 844  
 421 GGAAGGAACCTGGGGGGCTGGGGGAGGATGCAACCGCGGAGGAGGAGGAGGTGTGCAGA 480  
 845 AGGATGCGCCCTCCCAACAGCCCTTGTAGTGTGTGAGCCAGCTTAGTGTCTTGTG 904  
 481 AGGATGCGCCCTCCCAACAGCCCTTGTAGTGTGTGAGCCAGCTTAGTGTCTTGTG 540  
 905 AATCAAGACCCCTGACTTTCAGAGCTTGGGACACCCAGCGCAGCGCAGCGCTCCAG 964  
 541 AATCAAGACCCCTGACTTTCAGAGCTTGGGACACCCAGCGCAGCGCAGCGCTCCAG 600



```
/clone="2410050022"  
/cell_type="ES cells"  
/clone_lib="RIKEN full-length enriched mouse cDNA library"  
225..1217  
misc_feature  
/note="1600025D17RIK PROTEIN (PUTATIVE RETINOIC  
ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE)  
(SPTR|OSDAU1, evidence: PASTV, 100%ID, 100%length,  
match=828)  
putative"
```

## ORIGIN

```
Query Match 73.2%; Score 1345.6; DB 4; Length 1586;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1362; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
  
Qy 420 TGCAAGGGTATGTCGGAGACCTTTGAGACGCTGCAACACCTAGTCCAAAGGGGTCAA 479  
Db 215 TGCGAAGGGTATGTCGGAGACCTTTGAGACGCTGCAACACCTAGTCCAAAGGGGTCAA 274  
  
Qy 480 GGTGTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCT 539  
Db 275 GGTGTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCT 334  
  
Qy 540 CAAGAGCAGTGTGACGCTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGTGACAG 599  
Db 335 CAAGAGCAGTGTGACGCTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGTGACAG 394  
  
Qy 600 GAACACACAGAGGAGGACCTGACTGAATTCCTCTGTGCCAACCAACGCTGCTGAAAGGGAAA 659  
Db 395 GAAACACAGAGGAGGACCTGACTGAATTCCTCTGTGCCAACCAACGCTGCTGAAAGGGAAA 454  
  
Qy 660 GGACACAGTGTGCTAGCAGAGCGGTGTGTGGGAAGAGGAGACATAGACCTCCCTGGG 719  
Db 455 GGACACAGTGTGCTAGCAGAGCGGTGTGTGGGAAGAGGAGACATAGACCTCCCTGGG 514  
  
Qy 720 AGGAGAGAAATCCAGAGAGAGCGCAGCGGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAA 779  
Db 515 AGGAGAGAAATCCAGAGAGAGCGCAGCGGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAA 574  
  
Qy 780 GCAGAGGAGAACTGGGGGGGCTGGGGGAGAGATGCCAACGCCAGGAGGAGGAGGTGT 839  
Db 575 GCAGAGGAGAGACTGGGGGGGCTGGGGGAGAGATGCCAACGCCAGGAGGAGGAGGTGT 634  
  
Qy 840 GCAGAGGAGCATCG-CCCTCCACACAGCCGCCCTGTGATGAGCTGTGAGCCCGAGCTTAGTG 898  
Db 635 GCAGAGGAGCATCGCCCTCCACACAGCCGCCCTGTGATGAGCTGTGAGCCCGAGCTTAGTG 694  
  
Qy 899 TCCTTGATCAAGACCCCTGACTTCAGAGCTTGGGACAGCAGCAGCGCAGCGCAGCGCAG 958  
Db 695 TCCTTGATCAAGACCCCTGACTTCAGAGCTTGGGACAGCAGCAGCGCAGCGCAGCGCAG 754  
  
Qy 959 CTCAGCAAGGACAGCTGCTGTCCAGCATCAGGTCTCCTCTGGCTGTGCCCTTTCC 1018  
Db 755 CTCAGCAAGGACAGCTGCTGTCCAGCATCAGGTCTCCTCTGGCTGTGCCCTTTCC 814  
  
Qy 1019 TTCCCTTGAACACACAGCAAGAGGTGGAAGATCTGGGTGTCTGGGAGACGCAACCCCAA 1078  
Db 815 TTCCCTTGAACACACAGCAAGAGGTGGAAGATCTGGGTGTCTGGGAGACGCAACCCCAA 874  
  
Qy 1079 GGAAGAGAGGAGGAGGAGGAGGAGGAGCTCTCTTCTACAGTCCCCCTCAGAGCTCC 1138  
Db 875 GGAAGAGAGGAGGAGGAGGAGGAGGAGCTCTCTTCTACAGTCCCCCTCAGAGCTCC 934  
  
Qy 1139 GGGGTCCACCCAGCATCCAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGAGCA 1198  
Db 935 GGGGTCCACCCAGCATCCAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGAGCA 994  
  
Qy 1199 TGAGGCACATAGATGCTCACCAGCGCCCTTTCAGCCAGGAGGAGTCTCCGTGACGCTC 1258  
Db 995 TGAGGCACATAGATGCTCACCAGCGCCCTTTCAGCCAGGAGGAGTCTCCGTGACGCTC 1054  
  
Qy 1259 AGCAGCAGGCTGCTCTTCTTCCACAGCATTCTCTTCTGCTGCTCTTCTGCTCGAT 1318
```

```
Db 1055 AGCAGCCAGGCGCTGCCTCTTCTCCACCAAGCATTCTCTTCTGCTGCTGCTGCTCGAT 1114  
Qy 1319 GGTAAATTCGAGAACTTCCTCAGGACAAACTCGGGTGTGGCACAAGAGGGGTGAGACCCAGA 1378  
Db 1115 GGTAAATTCGAGAACTTCCTCAGGACAAACTCGGGTGTGGCACAAGAGGGGTGAGACCCAGA 1174  
Qy 1379 GCCAGAGCCAGCCAGAGACTGCGAGAGGACCTGACCTAAACCCCTTGGAAAGCCAA 1438  
Db 1175 GCCAGAGCCAGCCAGAGACTGCGAGAGGACCTGACCTAAACCCCTTGGAAAGCCAA 1234  
Qy 1439 TCTCAGTTCCCGTGTCCACCCACTCTCTCTGAGGAGCGCTCATGCTCTGCCAGCCCTT 1498  
Db 1235 TCTCAGTTCCCGTGTCCACCCACTCTCTCTGAGGAGCGCTCATGCTCTGCCAGCCCTT 1294  
Qy 1499 CTCCAGGGCTACAGAGTAAACACCTTTTGGCCCTTTCGGTTGGTTGGTCTGCTGCTCAT 1558  
Db 1295 CTCCAGGGCTACAGAGTAAACACCTTTTGGCCCTTTCGGTTGGTTGGTCTGCTGCTCAT 1354  
Qy 1559 CAGCTCCAGAGTGTCCCTCATCGATCTTTTGGCCCTTTCGGCCCAATCCAGGGCT 1618  
Db 1355 CAGCTCCAGAGTGTCCCTCATCGATCTTTTGGCCCTTTCGGCCCAATCCAGGGCT 1414  
Qy 1619 GGAAGGCCATCACCATTGAGGCTTAACTGTCTAGTACTAGGAGGTGCTGGAGCG 1678  
Db 1415 GGAAGGCCATCACCATTGAGGCTTAACTGTCTAGTACTAGGAGGTGCTGGAGCG 1474  
Qy 1679 CCCGGGTGGTTGGGTAAATCACTCACTGGCTCTCAAGCTTCTTAACTGAGCCCT 1738  
Db 1475 CCCGGGTGGTTGGGTAAATCACTCACTGGCTCTCAAGCTTCTTAACTGAGCCCT 1534  
Qy 1739 TAATACAGTCTCTCTGTTGGTGACTCCACGCCCCACACACACACCAT 1790  
Db 1535 TAATACAGTCTCTCTGTTGGTGACTCCACGCCCCACACACACACCAT 1586  
  
RESULT 9  
CX166520  
LOCUS HESC2_41_D10_g1_A035 NIH_MGC_258 Homo sapiens cDNA clone  
DEFINITION IMAGE:7470381.5', mRNA sequence.  
ACCESSION CX166520  
VERSION CX166520.1 GI:56796600  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 766)  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: BresaGen, Inc.  
cDNA Library Preparation: Express Genomics, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM15772 Row: g column: 19  
Seq primer: JENREV (CAGGAACAGCTATGACC)  
High quality sequence stop: 766.  
Location/Qualifiers  
1..766  
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
  
FEATURES  
source
```



/clone="IMAGE:7470381"  
/sex="male"  
/tissue\_type="human embryonic stem cells differentiated to  
an early endodermal cell type"  
/cell\_type="human embryonic stem cells"  
/cell\_line="BG01"  
/lab\_hosts="DH10B-T1 phage-resistant E. coli"  
/clone\_lib="NIH\_MGC\_258"  
/note="vector: pExpress-1; Site 1: NotI; Site 2: EcoRV;  
RNA obtained from human embryonic stem cells isolated from  
the inner cell mass of blastocyst stage embryos and  
differentiated to an early endodermal cell type. Cell line  
id and NIH Registry designation is BG01. Positive for  
GATM4, Mixl1, Mx1, HNP4alpha expression; negative for APP  
expression. Passage number 40. cDNA primed using oligo-dT  
primer: 5'-pGACTAGTCTAGATCGGAGCGGCCCT(T)25-3' and  
cloned into the EcoRV/NotI sites of pExpress-1. This  
primary library is non-normalized (normalized primary  
library is NIH MGC 259). It was constructed by Express  
Genomics (Frederick, MD). Sequence ends have been trimmed  
to exclude vector and regions below Phred quality 16.  
Three-prime sequences are presented as their reverse  
complement and have been trimmed to exclude polyA. Note:  
this is a Mammalian Gene Collection library."

ORIGIN

Query Match 41.7%; Score 766; DB 8; Length 766;  
Best Local Similarity 100.0%; Pred. No. 7.7e-179;  
Matches 766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

36 CTCTGGGTCCGGCTGGGCGATGGAGTCCATCTGAGCTCGCGCCGCTCCCTCTATT 95  
Db 1 CTCTGGGTCCGGCTGGGCGATGGAGTCCATCTGAGCTCGCGCCGCTCCCTCTATT 60

96 TCCTTTGTGTGTCTGCTCCGCTGCTCTCTCTCTGCCCCGAAGCTAGGCCGAGTCC 155  
Db 61 TCCTTTGTGTGTCTGCTCCGCTGCTCTCTCTCTGCCCCGAAGCTAGGCCGAGTCC 120

156 CGCGGGGCTGAGGAGACCGACTGGGTGCGATTGCCCCGAAATGCGAGTGTGCAAGTA 215  
Db 121 CGCGGGGCTGAGGAGACCGACTGGGTGCGATTGCCACGAAATGCGAAGTGTGCAAGTA 180

216 TGTGTCTGTGAGCTGAAGTGGCTTTTGAGAAACGGGAAGCAAGGAAGTGA 275  
Db 181 TGTGTCTGTGAGCTGAAGTGGCTTTTGAGAAACGGGAAGCAAGGAAGTGA 240

276 CACCGGCTATGGCATCTCGACGGGAGGGCTCTGGAGTCAAGTACACCAAGTCGACTT 335  
Db 241 CACCGGCTATGGCATCTCGACGGGAGGGCTCTGGAGTCAAGTACACCAAGTCGACTT 300

336 ACGGTTAATTGAAGTCACTGAGACCATTTGCAAGAGGCTTCTGACTACAGCCTGCACAA 395  
Db 301 ACGGTTAATTGAAGTCACTGAGACCATTTGCAAGAGGCTTCTGACTACAGCCTGCACAA 360

396 GGAGGAGTGCAGCAACCGGTTTGCCAAAGGATGTCGAGACCTTTGAGACGCTGCA 455  
Db 361 GGAGGAGTGCAGCAACCGGTTTGCCAAAGGATGTCGAGACCTTTGAGACGCTGCA 420

456 CAACCTAGTCCACAAAGGGGTCAAGTGGTGATGGATATCCCTATGAGCTGTGGAAACA 515  
Db 421 CAACCTAGTCCACAAAGGGGTCAAGTGGTGATGGATATCCCTATGAGCTGTGGAAACA 480

516 GACCTCAGCAGAGTGGTGCCTCAAGAACAGTGTGACGTGCTGGTGGGAAGATTGA 575  
Db 481 GACCTCAGCAGAGTGGTGCCTCAAGAACAGTGTGACGTGCTGGTGGGAAGATTGA 540

576 AGAGTGATTGGAGTGGTACAGGAACCAAGGAGGAACCTGACTGAATTCCTCTG 635  
Db 541 AGAGTGATTGGAGTGGTACAGGAACCAAGGAGGAACCTGACTGAATTCCTCTG 600

636 TGCCAAACCACTGCTGAAGGGAAGGACACAGATTGCTCTAGCAGAGCGGTGCTTGGCAA 695  
Db 601 TGCCAAACCACTGCTGAAGGGAAGGACACAGATTGCTCTAGCAGAGCGGTGCTTGGCAA 660

696 GAAGGGGACATAGCCTCCCTGGGAGGGAAGAAATCAAGAAGCGCAGCGAGTCAA 755  
Db 661 GAAGGGGACATAGCCTCCCTGGGAGGGAAGAAATCAAGAAGCGCAGCGAGTCAA 720

756 GGGCTCTCCAGTGGCAGCAGCAGCAGAGAGAGAACTGGGGGC 801  
Db 721 GGGCTCTCCAGTGGCAGCAGCAGCAGAGAGAGAACTGGGGGC 766

RESULT 10  
BG974240 766 bp mRNA linear EST 12-JUN-2001  
LOCUS 602843972F1 NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:4979944 5',  
DEFINITION mRNA sequence.  
ACCESSION BG974240  
VERSION BG974240.1 GI:14361877  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 766)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-@email.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth  
Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone description: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10978 row: 0 column: 17  
High quality sequence stop: 761.  
Location/Qualifiers  
1. 766  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4979944"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Mam4"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furth,  
NIH Reference for transgenic model: Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."

ORIGIN

Query Match 41.4%; Score 761.2; DB 2; Length 766;  
Best Local Similarity 99.6%; Pred. No. 1.2e-177;  
Matches 763; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

658 AAGGACACGAGTTCCCTAGCAGAGCGGTGGTCTCGCAGAGAGGGGGCATAGCCTCCCTG 717  
Db 1 AAGGACACGAGTTCCCTAGCAGAGCGGTGGTCTCGCAGAGAGGGGGCATAGCCTCCCTG 60

718 GGAGCGAAGAATCAAGAAGAACGCGACGCGAGTCAAGGGCTCTCCAGTGGCAGCAGC 777  
Db 61 GGAGCGAAGAATCAAGAAGAACGCGACGCGAGTCAAGGGCTCTCCAGTGGCAGCAGC 120

778 AAGCAGAGGAAGAACTGGGGGGCTGGGGGAGGATGCCAAGCGCCGAGGAGGAGGCT 837  
Db 121 AAGCAGAGGAAGAACTGGGGGGCTGGGGGAGGATGCCAAGCGCCGAGGAGGAGGCT 180



```
QY 838 GTGCAGAGGCAATCCCTCCCTCCCAACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGT 897
Db 181 GTGCAGAGGCAATCCCTCCCTCCCAACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGT 240
QY 898 GTCTCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGACGACACAGCGGAGCGCA 957
Db 241 GTCTCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGACGACACAGCGGAGCGCA 300
QY 958 GCTCAGCAAGGACAGCTGCTGTCCAGCATCAGGTCTCTCTCTTGGGTGTGCCCCCTTTC 1017
Db 301 GCTCAGCAAGGACAGCTGCTGTCCAGCATCAGGTCTCTCTCTTGGGTGTGCCCCCTTTC 360
QY 1018 TTCTCCCTTGAACAAACAGCAAGGTTGGAAGATCTGGGGTGTGGGAGACGGCACCCCAA 1077
Db 361 TTCTCCCTTGAACAAACAGCAAGGTTGGAAGATCTGGGGTGTGGGAGACGGCACCCCAA 420
QY 1078 AGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1137
Db 421 AGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
QY 1138 CGGGTCCACCCAGCATCCCAAGGCTGAGATCCAGGCTCCTGACATGGAAGCTGAAGAGC 1197
Db 481 CGGGTCCACCCAGCATCCCAAGGCTGAGATCCAGGCTCCTGACATGGAAGCTGAAGAGC 540
QY 1198 ATGAGGCAATAGATGCTCACAGGCCCCCTTCCAGCAGGAAGGATCCGCTGAGGCT 1257
Db 541 ATGAGGCAATAGATGCTCACAGGCCCCCTTCCAGCAGGAAGGATCCGCTGAGGCT 600
QY 1258 CAGCAGCAGGCTGCTCTCTCTTCCACCAAGCATCTCTCTCTGCTGCTTGTTCGGA 1317
Db 601 CAGCAGCAGGCTGCTCTCTCTTCCACCAAGCATCTCTCTCTGCTGCTTGTTCGGA 660
QY 1318 TGGTAAATTCAGAACTTCCAGGACAACTCGGGTGTGGCAAAAGGGCTGGAGCCAG 1377
Db 661 TGGTAAATTCAGAACTTCCAGGACAACTCGGGTGTGGCAAAAGGGCTGGAGCCAG 720
QY 1378 AGCCAGAGCCAGCCAGAGACTGCGAGAGGAGGACCTGACCTAAC 1423
Db 721 AGCCAGAGCCAGCCAGAGACTGCGAGAGGAGGACCTGACCTAAC 766
```

```
RESULT 11
LOCUS BQ922065 926 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT 8932400 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6468664
5' mRNA sequence.
ACCESSION BQ922065
VERSION BQ922065.1 GI:22337096
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 926)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13996 row: i column: 17
High quality sequence stop: 632.
Location/Qualifiers
1..926
/organism="Mus musculus"
```

RESULT 12  
B1104487  
LOCUS

B1104487 793 bp mRNA linear EST 26-JUN-2001

```
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6468664"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_94"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
```

## ORIGIN

```
Query Match 40.9%; Score 751.4; DB 5; Length 926;
Best Local Similarity 99.2%; Pred. No. 3.4e-175;
Matches 755; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 541 AAGAAAGCAGTGTGACGTCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACATGTTACAGG 600
Db 1 AAGAAAGCAGTGTGACGTCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACATGTTACAGG 60

QY 601 AACCAACAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAAGGGAAG 660
Db 61 AACCAACAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAAGGGAAG 120

QY 661 GACACGATTTGCTAGCAGAGCGGTGTCTGGCAAGAGAGGGGACATAGCTCTCCCTGGGA 720
Db 121 GACACGATTTGCTAGCAGAGCGGTGTCTGGCAAGAGAGGGGACATAGCTCTCCCTGGGA 180

QY 721 GGGGAAGAAATCCAAAGAAAGCGCAGCGAGTCAAGAGGCTCTCCAGTGGCAGCAGCAAG 780
Db 181 GGGGAAGAAATCCAAAGAAAGCGCAGCGAGTCAAGAGGCTCTCCAGTGGCAGCAGCAAG 240

QY 781 CAGAGGAAGAACTGGGGGGCTTGGGGGAGGATGCCAACCGCGAGGAGGAGGAGGTGTG 840
Db 241 CAGAGGAAGAACTGGGGGGCTTGGGGGAGGATGCCAACCGCGAGGAGGAGGAGGTGTG 300

QY 841 CAGAGGCAATCGCCCCCTCCCAACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGTGTG 900
Db 301 CAGAGGCAATCGCCCCCTCCCAACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGTGTG 360

QY 901 CTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACACCCACAGCCGACGCGCAGCT 960
Db 361 CTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACACCCACAGCCGACGCGCAGCT 420

QY 961 CCAGCAAGGACAGCTGTCTGTCAGCATCAGGTCTCTCTCTTGGCTGTGCCCCCTTTCTTT 1020
Db 421 CCAGCAAGGACAGCTGTCTGTCAGCATCAGGTCTCTCTCTTGGCTGTGCCCCCTTTCTTT 480

QY 1021 CCCTTGAAACACAGCAAGAGGTGGAAGATCTGGGGTGTGGGAGACGGCACCCCAAGG 1080
Db 481 CCCTTGAAACACAGCAAGAGGTGGAAGATCTGGGGTGTGGGAGACGGCACCCCAAGG 540

QY 1081 GAAGAGGAGGAGGAGGAGGAGGAGGAGCTCTCTTCTACAGTCCCTCCCTCAGGAGCTCCGG 1140
Db 541 GAAGAGGAGGAGGAGGAGGAGGAGGAGCTCTCTTCTACAGTCCCTCCCTCAGGAGCTCCGG 600

QY 1141 GGTCCACCCAGCATCCCGAGGCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCATG 1200
Db 601 GGTCCACCCAGCATCCCGAGGCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCATG 660

QY 1201 AGGCACATAAGATGCTACACAGCGCCCCCTTTCCAGCAGGAAGGATCTCGTGCAGCTCAG 1260
Db 661 AGGCACATAAGATGCTACACAGCGCCCCCTTTCCAGCAGGAAGGATCTCGTGCAGCTCAG 720

QY 1261 CAGCAGGCGCTGCTCTCTCTTCCACCAAGCATTTCTTCT 1301
Db 721 CAGCAGGCGCTGCTCTCTTCTTCCACCAAGCATTTCTTCT 761
```

DEFINITION 60289091F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:5035030 5',  
mRNA sequence.  
ACCESSION B1104487  
VERSION B1104487.1 GI:14555380  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
1 (bases 1 to 793)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11097 row: f column: 23  
High quality sequence stop: 784.  
FEATURES  
Location/Qualifiers  
1..793  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5035030"  
/tissue\_type="spontaneous tumor, metastatic to mammary."  
Stem cell origin."  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Lu29"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
ORIGIN  
Query Match 40.4%; Score 743.8; DB 2; Length 793;  
Best Local Similarity 98.7%; Pred No. 2.5e-173;  
Matches 781; Conservative 0; Mismatches 7; Indels 3; Gaps 3;  
Qy 9 GGAGGAAGCGCGCGGTCGCTCTGCTTGGGTCCGGTCCGATGGAGTCCATGTC 68  
Db 6 GGAGGAAGCGCGCGGTCGCTCTGCTCCGGTCCGCTGGGCGATGGAGTCCATGTC 65  
Qy 69 TGAGTCCGCGCGCGCTGCTCTTATTTCTTGTGCTGCTGCTCCGCTGCTGCTCCT 128  
Db 66 TGAGTCCGCGCGCGCTGCTCTTATTTCTTGTGCTGCTGCTCCGCTGCTGCTCCT 125  
Qy 129 TCCTGCCCGGAAGCTAGGCGCGAGTCCCGCGGCGCTGAGGAGCCGACTGGTCCGATT 188  
Db 126 TCCTGCCCGGAAGCTAGGCGCGAGTACC CGCGCGGCTGAGGAGCCGACTGGTCCGATT 185  
Qy 189 GCCCAGCAATGCCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTGAGGA 248  
Db 186 GCCCAGCAATGCCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGC-TTTGAGGA 244  
Qy 249 AACGGGAAGACCAAGGAAGTATTTGACACCGGCTATGGCATCTTGGACGGGAAGGGCTC 308  
Db 245 AACAGGAAGACCAAGGAAGTATTTGACACCGGCTATGGCATCTTGGACGGGAAGGGCTC 304  
Qy 309 TGGAGTCAAGTACACCAAGTCCGACTTACGTTTAACTGAAGTCACTGAGACCATTTGCAA 368  
Db 305 TGGAGTCAAGTACACCAAGTCCGACTTACGTTTAACTGAAGTCACTGAGACCATTTGCAA 364  
Qy 369 GAGGCTTCTGGACTACAGCTGCACCAAGGAGGAGGACTGGCGCAACCGGTTTGCAAGGG 428  
Db 365 GAGGCTTCTGGACTACAGCTGCACCAAGGAGGAGGACTGGCGCAACCGGTTTGCAAGGG 424

Qy 429 TATGTCGGAGACCTTTTCAGAGCGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGGTGAT 488  
Db 425 TATGTCGGAGACCTTTTCAGAGCGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGGTGAT 484  
Qy 489 GGATATCCCTTCATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCA 548  
Db 485 GGATATCCCTTCATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCA 544  
Qy 549 GTGTGAGCTGCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCA 608  
Db 545 GTGTGAGCTGCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCA 604  
Qy 609 GGAGGAAGACCTGACTCAATTTCTCTGTGCAACCACTGCTGTAAGGGAAGGACACGAG 668  
Db 605 GGAGGAAGACCTGACTCAATTTCTCTGTGCAACCACTGCTGTAAGGGAAGGACACGAG 664  
Qy 669 TTGCTTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCTCCCTGGGAGGAAGAA 728  
Db 665 TTGCTTAGCAGAGCGGTGGTCTGGC-AGAAAGGGGACATAGCTCCCTGGGAGGAAGAA 723  
Qy 729 ATCCAAAGAGCGGAGCGAGGTCAAGGGTCTCTCAGTGGCAGCAGCAGCAGAGGAA 788  
Db 724 ATCCAAAGAGAGCGCAGCGGGGTCAA-GGCTCTCCAGTGGCAGCAGCAGCAGAGGAA 782  
Qy 789 GGAACCTGGGG 799  
Db 783 GGAACCTGGGG 793  
RESULT 13  
CB203728  
LOCUS  
DEFINITION AGENCOURT\_11288067 NIH\_MGC\_135 Mus musculus cDNA clone  
IMAGE:30142422 5', mRNA sequence.  
CB203728  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM0051 row: k column: 07  
High quality sequence stop: 609.  
Location/Qualifiers  
1..937  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30142422"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_135"  
/note="Vector: pCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized full-length enriched library from pooled mouse  
embryonic limb, maxilla and mandible, day 12.5, 13.5,  
14.5, and 15.5 (size selected for the 0.5-1 kb fragments)  
Cloned directionally, priming method: Oligo-dT. cDNA  
enrichment: 51k bp, Average insert size 1.6k bp.  
Normalization (Cot value): 7.5 kb. Priming sequence:



1603 CCCAATCCAGGGCTGGAAGCCCATCACCATCATTTGAGGCTTAACCTGTCAGTTACTA 1662  
 541 CCCAATCCAGGGCTGGAAGCCCATCACCATCATTTGAGGCTTAACCTGTCAGTTACTA 600  
 1663 GGAGGTGCTGGAGCGCCCGGGTGTGTTGGGTAATCACTACTGCTCTCAGCCTTC 1722  
 601 NGAGGTGCTGGAGCGCCCGGGTGTGTTGGGTAATCACTACTGCTCTCAGCCTTC 660  
 1723 TAACACTGCAGCCCTTAATACAGTTCTTCTGTTGGTGAATCCACGCCCCCACA 1782  
 661 TAACACTGCAGCCCTTAATACAGTTCTTCTGTTGGTGAATCCACGCCCCCACA 720  
 1783 CACACCAATAAATT-ATTTCGATGCTGTTTCATAACTGTAAAAAAA 1828  
 721 CACACCAATAAATTAAATTCGATGCTGTTTCAAAAAAAA 767

RESULT 15  
 CF553710  
 LOCUS  
 DEFINITION AGENCOURT 15614073 Organ of Corti Mus musculus cDNA clone  
 IMAGE:30519414 5', mRNA sequence.  
 ACCESSION CF553710  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 892)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Daniela S. Gerhardt, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Bechara Kachar and Celine Pompeia, NIDCD  
 cDNA Library Preparation: Bechara Kachar, Celine Pompeia, NIDCD  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: NDAM590 row: 0 column: 07  
 High quality sequence stop: 579.  
 Location/Qualifiers  
 1..892  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30519414"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="Organ of Corti"  
 Note="Organ: Organ of Corti/Bar; Vector: pBluescriptSK-;  
 Site 1: EcoRI; Site 2: XhoI; The organ of Corti (OC) was  
 fine dissected from a total of 386 OC as follows: 102  
 samples from post-natal (P) day 5; 72 from P6; 60 from P7;  
 46 from P8; 18 from P9; 20 from P10; 14 from P12 and 24  
 from P13. Total RNA was extracted using the micro  
 Fasttrack kit (catalog # Ki593-02; Invitrogen, Carlsbad,  
 CA), according to manufacturer's instructions. Reverse  
 transcription and library construction were carried out  
 with the Uni-Zap XR vector kit (catalog # 237211,  
 Stratagene) and Uni-Zap XR Gigapack III Gold Cloning kit  
 (catalog # 237612), both from Stratagene (La Jolla, CA,  
 USA), according to manufacturer's instructions. The  
 frequency distribution of the library is as follows: 72%  
 of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and  
 0.1% 51-150. As to gene function, 45% of genes are present

FEATURES  
 source  
 1..892  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30519414"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="Organ of Corti"  
 Note="Organ: Organ of Corti/Bar; Vector: pBluescriptSK-;  
 Site 1: EcoRI; Site 2: XhoI; The organ of Corti (OC) was  
 fine dissected from a total of 386 OC as follows: 102  
 samples from post-natal (P) day 5; 72 from P6; 60 from P7;  
 46 from P8; 18 from P9; 20 from P10; 14 from P12 and 24  
 from P13. Total RNA was extracted using the micro  
 Fasttrack kit (catalog # Ki593-02; Invitrogen, Carlsbad,  
 CA), according to manufacturer's instructions. Reverse  
 transcription and library construction were carried out  
 with the Uni-Zap XR vector kit (catalog # 237211,  
 Stratagene) and Uni-Zap XR Gigapack III Gold Cloning kit  
 (catalog # 237612), both from Stratagene (La Jolla, CA,  
 USA), according to manufacturer's instructions. The  
 frequency distribution of the library is as follows: 72%  
 of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and  
 0.1% 51-150. As to gene function, 45% of genes are present

in GenBank and have know function; 23% have hits in  
 GenBank, but do not have assigned function; 12% are  
 uncharacterized ESTs and 20% are unidentified. Library  
 created in the laboratory of M. Brownstein (NIMH, NIH). A  
 complete library description can be found at "

ORIGIN  
 Query Match 39.9%; Score 733.6; DB 6; Length 892;  
 Best Local Similarity 97.9%; Pred. No. 8.8e-171;  
 Matches 753; Conservative 0; Mismatches 15; Indels 1; Gaps 1;  
 Qy 47 GCTGGCCATCGAGTCCATGTCGAGTCGCGCCCGCTCTTATTCCTTTGCTGC 106  
 Db 7 GGTGGCCATCGAGTCCATGTCGAGTCGCGCCCGCTCTTATTCCTTTGCTGC 66  
 Qy 107 TGCTGCTTCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 166  
 Db 67 TGCTGCTTCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 126  
 Qy 167 AGGAGACCGACTGGGTGCGATTGGCCAGCAAAATCGGAAGTGTGCAAGTATGTTGCTG 226  
 Db 127 AGGAGACCGACTGGGTGCGATTGGCCAGCAAAATCGGAAGTGTGCAAGTATGTTGCTG 186  
 Qy 227 AGCTGAAGTCGGCTTTTGAAGAAACGGGAAAGCAAGGAAGTGAATTCACACCGGCTATG 286  
 Db 187 AGCTGAAGTCGGCTTTTGAAGAAACGGGAAAGCAAGGAAGTGAATTCACACCGGCTATG 246  
 Qy 287 GCATCTCTGAGCGGGAAGGGCTCTGGAGTCAAGTACCAAGTCCGAGCTTACGGTTAATTG 346  
 Db 247 GCATCTCTGAGCGGGAAGGGCTCTGGAGTCAAGTACCAAGTCCGAGCTTACGGTTAATTG 306  
 Qy 347 AAGTCACTGAGACCAATTTGCAAGAGGGCTTCTGGACTACAGCTTCACAAAGAGAGGACTG 406  
 Db 307 AAGTCACTGAGACCAATTTGCAAGAGGGCTTCTGGACTACAGCTTCACAAAGAGAGGACTG 366  
 Qy 407 GCAGCAACCGGTTTGCACAGGGTATGTCGAGACCTTTGAGACCGCTGCAACCTAGTTC 466  
 Db 367 GCAGCAACCGGTTTGCACAGGGTATGTCGAGACCTTTGAGACCGCTGCAACCTAGTTC 426  
 Qy 467 ACAAGGGGTCNAGTGGTGGTATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAG 526  
 Db 427 ACAAGGGGTCNAGTGGTGGTATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAG 486  
 Qy 527 AGGTGGCTGACCTCAAGAAGCAGTGTGACGCTGCTGGTGGAGAGTTCGAGAGGTGATTG 586  
 Db 487 AGGTGGCTGACCTCAAGAAGCAGTGTGACGCTGCTGGTGGAGAGTTCGAGAGGTGATTG 546  
 Qy 587 AGGACTGGTACAGGAACCAACGAGGAGAACCTGACTGAATTCCTCTGTGCTCCCAACG 646  
 Db 547 AGGACTGGTACAGGAACCAACGAGGAGAACCTGACTGAATTCCTCTGTGCTCCCAACG 606  
 Qy 647 TGCTGAAGGGAAGGACACGAGTTGCTTACGAGCGGTGCTGTCGCAAGAGGGGGACA 706  
 Db 607 TGCTGAAGGGAAGGACACGAGTTGCTTACGAGCGGTGCTGTCGCAAGAGGGGGACA 666  
 Qy 707 TAGC-CTCCCTGGAGGGAAGAAATCCAAAGAAAGCGGCGAGCGAGTCAAGGGCTCTCTCC 765  
 Db 667 TAGC-CTCCCTGGAGGGAAGAAATCCAAAGAAAGCGGCGAGCGAGTCAAGGGCTCTCTCC 726  
 Qy 766 AGTGGCAGCAAGCAAGCAGAGGAAGAACTGGGGGGCTCTGGGGGAGGATG 814  
 Db 727 AGTGGCAGCAAGCAAGCAGAGGAAGAACTGGGGGGCTCTGGGGGAGG 775

Search completed: December 30, 2005, 03:38:20  
 Job time : 6737 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:03:20 ; Search time 9079 Seconds  
(without alignments)  
11513.945 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggacgaggggaggaagcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

1: gb ba:\*  
2: gb in:\*  
3: gb env:\*  
4: gb om:\*  
5: gb ov:\*  
6: gb pat:\*  
7: gb ph:\*  
8: gb pr:\*  
9: gb ro:\*  
10: gb sts:\*  
11: gb sy:\*  
12: gb un:\*  
13: gb vl:\*  
14: gb htg:\*  
15: gb pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1829.4	99.5	1882	9 BC013549	BC013549 Mus muscu
2	1784.2	97.0	1909	9 AF361644	AF361644 Mus muscu
3	1546.4	84.1	1655	9 AK220209	AK220209 Mus muscu
4	691.6	37.6	1413	8 BC008898	BC008898 Homo sapi
5	691.6	37.6	1413	8 BC008961	BC008961 Homo sapi
6	691.6	37.6	1675	6 AX464002	AX464002 Sequence
7	691.6	37.6	1675	8 AY358960	AY358960 Homo sapi
8	691.6	37.6	1720	8 BC004423	BC004423 Homo sapi
9	502.8	27.3	958	6 CQ716533	CQ716533 Sequence
10	491.2	26.7	962	8 HS080744	HS080744 Homo sapien
11	484	26.3	246404	14 AC131887	AC131887 Rattus no
12	467	25.4	1201	6 AR379595	AR379595 Sequence
13	458.6	24.9	1534	6 AR379729	AR379729 Sequence
14	457.8	24.9	1277	8 BC008133	BC008133 Homo sapi
15	457.8	24.9	1593	8 BC011767	BC011767 Homo sapi
16	446	24.3	1657	6 BD194905	BD194905 86 human
17	446	24.3	1657	6 CQ855233	CQ855233 Sequence
18	416	22.6	222256	14 AC134482	AC134482 Rattus no

19	412.2	22.4	580	6 AX198865	AX198865 Sequence
20	412.2	22.4	580	6 AX209392	AX209392 Sequence
21	359.8	19.6	2623	8 AK124196	AK124196 Homo sapi
22	339.4	18.5	917	5 CR524291	CR524291 Gallus ga
23	331.2	18.0	904	8 BC022093	BC022093 Homo sapi
24	309.2	16.8	1503	5 BC088872	BC088872 Xenopus t
25	300	16.3	758	5 BC083297	BC083297 Danio rer
26	298	16.2	2763	5 BC073680	BC073680 Xenopus l
27	284	15.4	612	8 AF161347	AF161347 Homo sapi
28	266.8	14.5	4469	8 AK090425	AK090425 Homo sapi
29	263.6	14.3	526	6 CQ921127	CQ921127 Sequence
c	186	10.1	89855	14 AC074220	AC074220 Mus muscu
31	172	9.4	1020	9 BC086647	BC086647 Mus muscu
32	172	9.4	1526	6 AX590165	AX590165 Sequence
33	172	9.4	1958	9 BC094672	BC094672 Mus muscu
34	160.2	8.7	744	6 AR142811	AR142811 Sequence
35	160.2	8.7	746	6 BD123856	BD123856 Secretary
36	160.2	8.7	746	6 CS051652	CS051652 Sequence
37	160.2	8.7	746	6 AX136616	AX136616 Sequence
38	160.2	8.7	747	6 BD006701	BD006701 Novel pol
39	160.2	8.7	1413	8 BC019903	BC019903 Homo sapi
40	160.2	8.7	1419	6 BD123666	BD123666 Secretary
41	160.2	8.7	1419	6 CS051459	CS051459 Sequence
42	160.2	8.7	1419	6 AX136423	AX136423 Sequence
43	160.2	8.7	1419	8 AK075537	AK075537 Homo sapi
44	160.2	8.7	1457	6 BD137191	BD137191 Mammalian
45	160.2	8.7	1462	6 BD227243	BD227243 Secreted

#### ALIGNMENTS

RESULT 1  
BC013549  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BC013549 1882 bp mRNA linear ROD 28-SEP-2004  
Mus musculus trinucleotide repeat containing 5, mRNA (cDNA clone MGC:19340 IMAGE:4222133), complete cds.

BC013549  
MGC.  
Mus musculus (house mouse)

BC013549  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 1882)  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, J.A., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Viall, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Touchman, J.W., Green, E.D., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Schmutz, J., Myers, R.M., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932

2 (bases 1 to 1882)  
Director MGC Project.  
Direct Submission

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [c9apbs-remail.nih.gov](mailto:c9apbs-remail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@axil.stanford.edu](mailto:mcd@axil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 25 Row: j Column: 9  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31982623.

FEATURES		Location/Qualifiers
source	1..1882	/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="FVB/N"
		/db_xref="taxon:10090"
		/clone="MGC:19340 IMAGE:4222133"
		/tissue_type="kidney, normal. 5 month old male mouse."
		/clone_lib="NCI CGAP_Kid14"
		/lab_host="DH10B"
		/note="Vector: pCMV-SPORT6"
	1..1882	
gene		/gene="Tnrc5"
		/note="synonyms: CAG4A, ERDA5"
		/db_xref="GeneID:72029"
		/db_xref="MGI:1919279"
	91..921	
		/gene="Tnrc5"
		/codon_start=1
		/product="trinucleotide repeat containing 5"
		/protein_id="AAH13549.1"
		/db_xref="GI:15488836"
CDS		/db_xref="GeneID:72029"
		/db_xref="MGI:1919279"
		/translation="MESSELPAPLLPPLLLPAPKLGSPAGAEFDWVR LPSCEVKYVAVELKSFETGKTEVIDTGYLGDKSGVKYKTSDLRIEVTET ICKLLDYSLHKERTGNSRFAKSETFETLHNLHVKGKVMVDIPYELNETSAEVA DLKQCDVLFEEFEVIEDWYRNHQBEDLTFELCANHVLKGDTSCLAERWSGKKGDI ASLGGKSKKRSRGVKGSSGSKQKELGLGSDANAEEBEGVQKASPLPHSPDEL "
ORIGIN		Query Match
		Best Local Similarity 99.5%; Score 1829.4; DB 9; Length 1882;
		Matches 1830; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	9	GGAGGAAGCGCGCGGCTCGCTCTGCTCGGGTCGGCTGGCCATGGAGTCCATGTC 68
Db	45	GGAGGAAGCGCGCGGCTCGCTCTGCTCGGGTCGGCTGGCCATGGAGTCCATGTC 104
Qy	69	TGAGCTCGCGCGCGCTGCTTATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 128
Db	105	TGAGCTCGCGCGCGCTGCTTATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 164
Qy	129	TCCTGCCCGAAGCTAGGCGCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTCGATT 188
Db	165	TCCTGCCCGAAGCTAGGCGCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTCGATT 224
Qy	189	GCCAGCAAAATCCGAAGTGTCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTGAGGA 248
Db	225	GCCAGCAAAATCCGAAGTGTCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTGAGGA 284
Qy	249	AACGGGAAGACCAAGGAAGTGATTGACACCGGCTATGGCATCTCGGACGGGAAGGCTC 308

285	AAACGGGAAGACCAAGGAAGTGATTGACACCGCTATGGCATCTCGGACGGGAAGGGCTC 344	Db
309	TGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATTGAAGTCACTGAGACCAATTGCAA 368	Qy
345	TGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATTGAAGTCACTGAGACCAATTGCAA 404	Db
369	GAGGCTTCTGACTACAGCCCTGCAAGAGGAGAGACTGGCAGCAACCGGTTTGCCAAAGGG 428	Qy
405	GAGGCTTCTGACTACAGCCCTGCAAGAGGAGAGACTGGCAGCAACCGGTTTGCCAAAGGG 464	Db
429	TATGTCGAGACCTTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAAAGGTGGTGAT 488	Qy
465	TATGTCGAGACCTTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAAAGGTGGTGAT 524	Db
489	GGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCA 548	Qy
525	GGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCA 584	Db
549	GTGTGACGTGCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCAACA 608	Qy
585	GTGTGACGTGCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCAACA 644	Db
609	GGAGGAAGACCTGACTGAAATTCCTCTGTGCAACCAACCTGCTGAAAGGGAAGGACACGAG 668	Qy
645	GGAGGAAGACCTGACTGAAATTCCTCTGTGCAACCAACCTGCTGAAAGGGAAGGACACGAG 704	Db
669	TTGCTTAGCAGAGCGGTGCTGTGCAAGAGAGGGGGAATAGCCTCCCTGGAGGGAAGAA 728	Qy
705	TTGCTTAGCAGAGCGGTGCTGTGCAAGAGAGGGGGAATAGCCTCCCTGGAGGGAAGAA 764	Db
729	ATCCAAGAAGACGCGAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAGCAGAGAGGAA 788	Qy
765	ATCCAAGAAGACGCGAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAGCAGAGAGGAA 824	Db
789	GGAACTCGGGGGCTCGGGGGAGTGCACACCGCGAGGAGGAGGAGGTGTGCAGAAAGGC 848	Qy
825	GGAACTCGGGGGCTCGGGGGAGTGCACACCGCGAGGAGGAGGAGGTGTGCAGAAAGGC 884	Db
849	ATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGTGTCTTGAATC 908	Qy
885	ATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGTGTCTTGAATC 944	Db
909	AAGACCCCTGACTTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCGCTCCAGCAAG 968	Qy
945	AAGACCCCTGACTTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCGCTCCAGCAAG 1004	Db
969	GACAGCTGCTGTCCAGCATCAGGTCTCTCCCTTTGGCTGTGCCCTTTTCTTCCCTTGAA 1028	Qy
1005	GACAGCTGCTGTCCAGCATCAGGTCTCTCCCTTTGGCTGTGCCCTTTTCTTCCCTTGAA 1064	Db
1029	CAACAGCAAGGTGGGAAGGATCTGGGGTGTCTGGGACCGCACCCCAAGGGGAAGAGGA 1088	Qy
1065	CAACAGCAAGGTGGGAAGGATCTGGGGTGTCTGGGACCGCACCCCAAGGGGAAGAGGA 1124	Db
1089	GGAGGACGAGAGGACGCTCTCTTTTACACAGTCCCTCCCTCAGGAGTCCGGGGTCCACC 1148	Qy
1125	GGAGGACGAGAGGACGCTCTCTTTTACACAGTCCCTCCCTCAGGAGTCCGGGGTCCACC 1184	Db
1149	CAGCATCCCGAGGCTGAGATCCAGGCTCCTGACATGGAAGCTGAAGAGCATGAGGCACAT 1208	Qy
1185	CAGCATCCCGAGGCTGAGATCCAGGCTCCTGACATGGAAGCTGAAGAGCATGAGGCACAT 1244	Db
1209	AAGATGCTACACGCGCCCCCTTACGAGGAGGACTCGTGCAGCTCAGCAGCCAGG 1268	Qy
1245	AAGATGCTACACGCGCCCCCTTACGAGGAGGACTCGTGCAGCTCAGCAGCCAGG 1304	Db
1269	CCTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGCTCTTCTGCTGATGGTAAATTCG 1328	Qy
1305	CCTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGCTCTTCTGCTGATGGTAAATTCG 1364	Db
1329	AGAACTTCCAGGACAAACTCGGGTGTGGCAAAAGGGGCTGGACCGCAGAGCCAGAGCCA 1388	Qy





Qy	909	AAGACCCCTGACTTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCTCCAGCAAG	968
Db	943	AAGACCCCTGACTTTCAGAGCTTGGGACACGACA - CGCAGCGCAGCGCAGCTCCAGCAAG	1001
Qy	969	GACAGCTGCTGTCCAGCATCAGGTCTCTCTCCCTTGGCTGTGCCCTTTCTTCCCTTGAA	1028
Db	1002	GACAGCTGCTGTCCAGCATCAGGTCTCTCTCCCTTGGCTGTGCCCTTTCTTCCCTTGAA	1061
Qy	1029	CACAGCAAGAGGTGGAAGATCTGGGGTGTCTGGGACGCGCACCCCAAAGGAAGAGGA	1088
Db	1062	CACAGCAAGAGGTGGAAGATCTGGGGTGTCTGGGACGCGCACCCCAAAGGAAGAGGA	1121
Qy	1089	GGAGGACGAAAGGCAGCTCTCTTTACACAGTCCCCCTCACAGCTCCGGGGTCCACC	1148
Db	1122	GGAGGACGAAAGGCAGCTCTCTTTACACAGTCCCCCTCACAGCTCCGGGGTCCACC	1181
Qy	1149	CAGCATCCCAGGCTGAGATCCAGGCTCTTGACATGAAGTGAAGCATGAGGCACAT	1208
Db	1182	CAGCATCCCAGGCTGAGATCCAGGCTCTTGACATGAAGTGAAGCATGAGGCACAT	1241
Qy	1209	AAGATGCTCACGAGCGCCCTTCAGCCAGGAGGACTCCGTGCGAGCTCAGCAGCCAGG	1268
Db	1242	AAGATGCTCACGAGCGCCCTTCAGCCAGGAGGACTCCGTGCGAGCTCAGCAGCCAGG	1301
Qy	1269	CCTGCTCTTCTCTTCCACCAAGCATCTCTTCTGCTGGTCTTGTGCGATGTAATTCG	1328
Db	1302	CCTGCTCTTCTCTTCCACCAAGCATCTCTTCTGCTGGTCTTGTGCGATGTAATTCG	1361
Qy	1329	AGAACTTCCAGGACAAACTCGGGTGTGGCAAAAGGGGCTGGAGCCGAGAGCCAGGCCA	1388
Db	1362	AGAACTTCCAGGACAAACTCGGGTGTGGCAAAAGGGGCTGGAGCCGAGAGCCAGGCCA	1421
Qy	1389	CGCAGAGACTGCAGAGAGGGCACTGACCTAACCCCTCGGAAGCCATCTCGAGTTC	1448
Db	1422	CGCAGAGACTGCAGAGAGGGCACTGACCTAACCCCTCGGAAGCCATCTCGAGTTC	1481
Qy	1449	CCGTGTCCACCCACTCTCTCTGAGAGCGCTCATGCTCTGCCAGGCCCTTCTCCAGGGC	1508
Db	1482	CCGTGTCCACCCACTCTCTCTGAGAGCGCTCATGCTCTGCCAGGCCCTTCTCCAGGGC	1541
Qy	1509	TACAGAGTAAACCTTTTGGCCTTTTCGGTTTGGTTCTGGGTCTCATCAGCTCCAG	1568
Db	1542	TACAGAGTAAACCTTTTGGCCTTTTCGGTTTGGTTCTGGGTCTCATCAGCTCCAG	1601
Qy	1569	AGTGTCCCTCATCGATCTTTTGTGCTTTGTCCCCCAATCCAGGGGCTGGAAGGCCAT	1628
Db	1602	AGTGTCCCTCATCGATCTTTTGTGCTTTGTCCCCCAATCCAGGGGCTGGAAGGCCAT	1661
Qy	1629	CACCATCATTTGGAGGCTTAACCTGTCAAGTTACTAGGAGGTGCTGGGAGCGCCCGGGTTG	1688
Db	1662	CACCATCATTTGNAGGNTTAACCTGTCAAGTTACTAGGAGGTGCTGGGAGCGCCCGGGTTG	1721
Qy	1689	GTTTGGGGTAATCACTCACTGGCTCTCAGGCTTCTAACTGCGAGCCCCCTTAATACAGTT	1748
Db	1722	GTTTGGGGTAATCACTCACTGGCTCTCAGGCTTCTAACTGCGAGCCCCCTTAATACAGTT	1781
Qy	1749	CCTTCTGTTGTGGTACTCCGACGCCCGCACACACACATAAAATTATTTGATGCTG	1808
Db	1782	CCTTCTGTTGTGGTACTCCGACGCCCGCACACACACATAAAATTATTTGATGCTG	1841
Qy	1809	TTTTCATAACTGTAA	1822
Db	1842	TTTTCATAACTGTAA	1855
RESULT 3			
AK220209			
LOCUS	AK220209	1655 bp	linear
DEFINITION	Mus musculus mRNA for mFLJ00338 protein.	mRNA	ROD 01-MAR-2005
ACCESSION	AK220209		
VERSION	AK220209.1		
KEYWORDS	FLI CDNA.		





Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: iRAIc Place: 2 Row: m Column: 12  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5729791.  
 Location/Qualifiers

```

FEATURES             source
     1..1413
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="MGC:1220 IMAGE:2959532"
     /tissue_type="Colon, adenocarcinoma"
     /clone_lib="NIH MGC 15"
     /lab_host="DH10B-R"
     /note="Vector: pOTB7"
     1..1413
     /gene="TNRC5"
     /note="synonyms: ERDA5, CAG4A"
     /db_xref="GeneID:10695"
     69..905
     /gene="TNRC5"
     /codon_start=1
     /product="trinucleotide repeat containing 5"
     /protein_id="AAH08898.1"
     /db_xref="GI:14286204"
     /db_xref="GeneID:10695"
     /translation="MDSNPBPASRCLLLPLLLLLLLLPAPELGPQAGAEENDWVR
     LPSKECVKYAVELSGAFETGKTEVIGYGLDQKAGSVKIKSDLRLEIVET
     ICKRLDYLHKEFTGSRNFAKMGSEFTELNLVHRGVKVMVDIPYELWNETSAEVA
     DLKQCDLVIEFEEDYRNHQEEDLTFELCANHLVKGKDTSCLAEQMSGKGGDT
     AALGGKSKKSSRAKAAGRRSSSKQKELGGLGEGDPSPEEDBGIQKASPLTHSPD
     EL"
    
```

ORIGIN

```

Query Match       37.6%; Score 691.6; DB 8; Length 1413;
Best Local Similarity 84.5%; Pred. No. 1.9e-158;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

Qy  5  CGAGGGAGGAGCGCGCGGTCGCTCTGCTGGTCCGGTGGCCATGAGTCCA 64
Db  19 CGCGGGAGGAGGAGAACCGCGGTCCTTTAGGGTCCGGCCCGCGGCCATGATCAA 78
Qy  65 TGTCTGAGCTCGCGCCCGCTCTTATTTCTTCTGCTGCTGCTGCTGCTGCTG 124
Db  79 TGCTTGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 138
Qy  125 TCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 184
Db  139 TGCTGCGCGCCCGGAGCTGGGCGCGAGCCAGGCGCGAGCTGAGGAGAACGACTGGGTT 198
Qy  185 GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAGTCCGCTTTG 244
Db  199 GCCTGCCAGCAATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAGTCCGCTTTG 258
Qy  245 AGGAAACCGGAAAGACCAAGGAGTGTGACACCGGCTATGGCATCTGACCGGAGG 304
Db  259 AGGAAACCGGAAAGACCAAGGAGTGTGACACCGGCTATGGCATCTGACCGAGG 318
Qy  305 GCTCTGGAGTCAAGTACACCAAGTCCGACTTACCGTTAAATTAAGTCACTGAGACCAATTT 364
Db  319 CTTCTGGAGTCAATATACCAAGTCCGACTTACCGTTAAATTAAGTCACTGAGACCAATTT 378
Qy  365 GCAAGAGCTTCTGAGCTACAGCTGCAAGGAGAGGAGTGGCAGCAACCGTTTGGCA 424
Db  379 GCAAGAGCTTCTGAGTATATGAGCTGCAAGGAGAGGAGGAGCAATCGATTTGGCA 438
Qy  425 AGGTATGCTGGAGACCTTTGAGAGCTGCAACAGTGTAGTCCACAAAGGGGTCAAGGTGG 484
Db  439 AGGCAATGTCAGAGACCTTTGAGACATTTACAAACCTGTTACAAAGGGGTCAAGGTGG 498
Qy  485 TGATGGATATCCCTATGAGCTGTGGACGAGACCTTCAGCAGAGGTGGCTGACCTCAAGA 544
Db  499 TGATGGACATCCCTATGAGCTGTGGACGAGACCTTCTGACAGAGGTGGCTGACCTCAAGA 558
    
```

```

Qy  545 AGCAGTCTGAGCTCTGGTGGAGCAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACC 604
Db  559 AGCAGTGTGATGCTCTGGTGGAGAGGTTTGAAGAGGTGATCGAGACTGGTACAGGAACC 618
Qy  605 ACCAGGAGGAGAGCCTGACTGAATTCCTCTGTGCAACACGCTGCTGAAGGAGGAAGGACA 664
Db  619 ACCAGGAGGAGACCTGACTGAATTCCTCTGTGCGCAACACGCTGCTGAAGGAGGAAGACA 678
Qy  665 CGAGTTCCTAGCAGAGCGGTGGTCTGCAAGAGAGGGGACATAGCCTCCCTCGGAGGGA 724
Db  679 CCAGTTCCTGCGCAGAGCAGTGGTCCGCAAGAGAGGGAGACACAGCTGCCCTCGGAGGGA 738
Qy  725 AGAATCCAGAAGAAGCGCAGCGGAGTCAAG-----GGCTCTCTCCAGTGGCAGCACA 778
Db  739 AGAAGTCAAGAAGAAGCAGCGGCGCAAGGAGGCGGCGGAGGAGTAGCAGCAGCA 798
Qy  779 AGCAGAGGAGGAACTGGGGGCGCTGGGGGAGGATGCCAACCGCAGGAGGAGGAGGTG 838
Db  799 AACAAAGGAGGAGCTGGGTGGCTTGAGGGGAGACCCCGAGCCCGGAGGAGATGAGGGCA 858
Qy  839 TGCAGAGGAGTCCCTCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db  859 TCCAGAGGAGTCCCTCTCTCACACAGCCCCCTGATGAGCTGTGAGCCCGAGCTTAGTG 918
Qy  899 TCCTTTGAATCAAGACCCCTGACTTCAGAGCTTTGGGA 934
Db  919 CCTCTGCTCGAGACCCCTGATTTTGAAGCTGAGGA 954

RESULT 5
LOCUS      BC008961      1413 bp      mRNA      linear      PRI 29-JUN-2004
DEFINITION Homo sapiens trinucleotide repeat containing 5, mRNA (cDNA clone
ACCESSION   BC008961
VERSION     BC008961.2 GI:33870152
KEYWORDS    MSC.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 1413)
            Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
            Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
            Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
            Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
            Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
            Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
            Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Tothiyuki, S.,
            Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
            Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
            McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
            Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
            Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
            Fahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Shevchenko, Y.,
            Sanchez, A., Whiting, M., Madan, A., Touchman, J.W., Green, E.D.,
            Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
            Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
            Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalish, D.E.,
            Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
            Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
            Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
            12477932
            2 (bases 1 to 1413)
            Strausberg, R.
            Direct Submission
            Submitted (29-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
    
```

319	Db	CCTCTGGAGTCAAAATACACCAAGTCGGACTTTGCGGTAAATCGAAGTCACTGAGACCATTT	378
365	Qy	GCAGAAGGCTTCTGGACTACAGCCTTCGACAAAGGAGAGGACTGGCGGACGAACCGGTTTGCCA	424
379	Db	GCAGAAGGCTCCTGGATTATAGCCTTCGACAAAGGAGAGGACCGGACGAATTCGATTTTGCCA	438
425	Qy	AGGTATGTTCGGAGACCTTTTGAGACGCTGCAACAACCTAGTCCACAAGAGGGGTCAAGGTGG	484
439	Db	AGGCATGTCAGAGACCTTTTGAGACATTACACAACCTGGTACACAAGAGGGGTCAAGGTGG	498
485	Qy	TGATGGATATCCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGA	544
499	Db	TGATGGACATCCCCTATGAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA	558
545	Qy	AGCATGTGAGCTGCTGGTGGAAAGATTGAAAGAGGTGATTGAGGACTGGTATACAGGAACC	604
559	Db	AGCAGTGTGATGTCTGGTGGAAAGATTGAGGAGGTGATCGAGGACTGGTACAGGAACC	618
605	Qy	ACCAGGAGGAAGACTGACTGAAATTCCTCTGTGCGCAACACGCTGCTGAAAGGAAAGGACA	664
619	Db	ACCAGGAGGAAGACTGACTGAAATTCCTCTGTGCGCAACACGCTGCTGAAAGGAAAGACA	678
665	Qy	CGAGTTCCTTAGCAGAGCGGTGGTCTGGCAAGAAGGGGGACATAGCCTCTCCCTGGGAGGA	724
679	Db	CCAGTTCCTTGGCAGACAGTGGTCTGGCAAGAAGGGGAGACACAGCTGCCCTGGGAGGA	738
725	Qy	AGAAATCCAAAGAGAAGCGCAGCGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA	778
739	Db	AGAAATCCAAAGAGAAGCAGCAGCGGCTCAAGGCGACGAGGGGAGGAGTATGACAGCAGCA	798
779	Qy	AGCAGAGGAAGGAACCTGGGGGCTGGGGGAGGATGCCAAGCCGAGGAGGAGGAGGGTG	838
799	Db	AACAAAGGAAGGAGCTGGGTGGCTTTGAGGAGAGCCCCAGGCCCCGAGGAGGATGAGGGCA	858
839	Qy	TGCAGAAGGCATCGCCCTCCCAACACAGCCCCCTTGATGACGTGTGAGCCCGAGCTTAGTG	898
859	Db	TCCNAGAGGCATCCCTCTCACAACAGCCCCCTTGATGAGCTCTGAGGCCCCAGCCGAT	918
899	Qy	TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA	934
919	Db	CCTCTGCTCTGAGACCCCTGATTTTGAGCTTGAGA	954

RESULT 6	AX464002	AX464002	Sequence	135 from Patent WO0140466.	1675 bp	DNA	linear	PAT 16-JUL-2000
LOCUS	AX464002	AX464002	DEFINITION					
ACCESSION	AX464002	AX464002	VERSION	1	GI:21899018			
KEYWORDS			SOURCE	Homo sapiens (human)				
ORGANISM			SOURCE	Homo sapiens				
				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
				Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;				
				Hominidae; Homo.				
REFERENCE			1					
AUTHORS				Baker, K.P., Beresini, M., Deforge, L., Desnovers, L., Filvaroff, E., Gao, W.Q., Gerritsen, M.E., Goddard, A., Godowski, P.J., Gurney, A.L., Sherwood, S., Smith, V., Stewart, T.A., Tumas, D., Watanabe, C.K., Wood, W.L. and Zhang, Z.				
TITLE				Secreted and transmembrane polypeptides and nucleic acids encoding				
				same				
JOURNAL				Patent: WO 0140466-A 135 07-JUN-2001;				
				Genentech Inc. (US)				
FEATURES				Location/Qualifiers				
source				1..1675				
				/organism="Homo sapiens"				
				/mol_type="unassigned DNA"				
				/db_xref="taxon:9606"				
ORIGIN								
				Query Match	37.6%	Score	691.6;	DB 6;
				Best Local Similarity	84.5%	Pred.No.	1.9e-158;	Length 1675;

Matches	791;	Conservative	0;	Mismatches	139;	Indels	6;	Gaps	1;
Qy	5	CGAGGGAGGAGCGCCCGCGGTCCGCTCTGCTCTGGGTCCGGTCCGGCCATGAGTCCA	64						
Db	287	CGCGGGAGGAGGAAACCGCCCGGTCCTTAGGGTCCGGGCCCGCGGCCCATGATTCAA	346						
Qy	65	TGCTCTGAGCTCGCGCCCGCGCTCTATTTCCTTTGCTGCTGCTGCTCCGCTGCTGC	124						
Db	347	TGCTCTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCCCTGCTGCTGCTGCTGCTGC	406						
Qy	125	TCCTTCTGCCCCGAAAGCTAGCGCCGAGTCCCGCGGGGCTGAGAGACCGACTGGGTGC	184						
Db	407	TGCTCGCGGCCCGAGCTGGGGCCGAGCCAGGCGGAGCTGAGGAGAAACGACTGGGTTT	466						
Qy	185	GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTG	244						
Db	467	GCCTGCCAGCAATGCGAAGTGTGTAATATGTTGCTGTGGAGCTGAAGTCAGCCCTTG	526						
Qy	245	AGGAAACGGGAAAGACCAAGGAAGTGATTGACACCGGCTATGGCATCTCGACCGGAAGG	304						
Db	527	AGGAAACCGGCAAGACCAAGAGGTGATTGGCAGCGGCTATGGCATCTTGACCAAGG	586						
Qy	305	GCTCTGGAGTCAAGTACACCAAGTCGGACTTACCGTTAATTGAAGTCACTGAGACCAATT	364						
Db	587	CCTCTGGAGTCAATACACCAAGTCGGACTTTCGGTTAATCGAAGTCACTGAGACCAATT	646						
Qy	365	GCAAGAGCTTCTGGACTACAGCTGCACAGGAGAGGACTGGCAGCAACCGGTTTGCCA	424						
Db	647	GCAAGAGCTCTGGATTATATACCTGTCACAAAGGAGAGACCGGCAAGCAATCGATTG	706						
Qy	425	AGGGTATGTCGGAGACCTTTCAGAGCGTGCACACCTAGTCCACAAAGGGGTCAAGGTGG	484						
Db	707	AGGGCATGTCAAGACCTTTGAGACATTACACACCTGGTATACACAAAGGGGTCAAGGTGG	766						
Qy	485	TGATGGATATCCCTATAGCTGTGGAAACGAGACCTCAGCAGAGTGGCTGACCTCAAGA	544						
Db	767	TGATGGACATCCCTATAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA	826						
Qy	545	AGCAGTGTGACGTGCTGTGTAAGAGTTTGAAGGTGATTGAGGACTGTTACAGGAACC	604						
Db	827	AGCAGTGTGATGTGCTGTGTAAGAGTTTGAAGAGGTGATCGAGGACTGGTACAGGAACC	886						
Qy	605	ACCAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACACAGCTGCTGAAGGGAAGGACA	664						
Db	887	ACCAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACACAGCTGCTGAAGGGAAGACA	946						
Qy	665	CGAGTTGCTTAGCAGAGCGGTGTTGTCGAAGAAGGGGGACATAGCCTCCCTGGGAGGGA	724						
Db	947	CCAGTTGCTTGCAGAGCAGTGGTCCGGCAAGAAGGGAGACACAGCTGCCCTGGGAGGGA	1006						
Qy	725	AGAAATCCAAGAAGCGCAGCGAGTCAAG-----GGCTCTCTCCAGTGGCAGCAGCA	778						
Db	1007	AGAAATCCAAGAAGACGACGAGCGGCGCAAGGCAGCAGCGGCGGAGGAGTAGCAGCAGCA	1066						
Qy	779	AGCAGAGGAAGGACTGGGGGCGCTGGGGGAGGATGCCAACCGCAGGAGGAGGAGGGTG	838						
Db	1067	AACAAGGAAGAGCTGGGTGGCTTTGAGGAGACCCCGAGGAGGATGAGGCA	1126						
Qy	839	TGCAGAGGCATCGCCCTCCACACACAGCCCTCTGATGAGTGTGAGCCCGAGCTTAGTG	898						
Db	1127	TCCAGAGGCATCCCTCTCACACACAGCCCTCTGATGAGTCTGTAGGCCACCCAGCAT	1186						
Qy	899	TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA	934						
Db	1187	CCTCTGCTCAGACCCCTGATTTTGAAGCTGAGGA	1222						
RESULT 7									
AY358960									
LOCUS	AY358960		1675 bp	mRNA	linear	PRI 03-OCT-2003			
DEFINITION	Homo sapiens clone DNA88004	CTG4A	(UNQ1934)	mRNA	complete cds.				
ACCESSION	AY358960								
VERSION	AY358960.1	GI:37183037							

KEYWORDS	FLI CDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 1675) Clark,H.F., Gurney,A.L., Abaya,E., Baker,K., Baldwin,D., Brush,J., Chen,J., Chow,B., Chui,C., Crowley,C., Currell,B., Deuel,B., Dowd,P., Eaton,D., Foster,J., Grimaldi,C., Gu,Q., Haas,P.E., Heldens,S., Huang,A., Kim,H.S., Klimowski,L., Jin,Y., Johnson,S., Lee,J., Lewis,L., Liao,D., Mark,M., Robbie,E., Sanchez,C., Schoenfeld,J., Seshagiri,S., Simmons,L., Singh,J., Smith,V., Stinson,J., Vagts,A., Vanden,R., Watanabe,C., Weand,D., Woods,K., Xie,M.H., Yansura,D., Yi,S., Yu,G., Yuan,J., Zhang,M., Zhang,Z., Goddard,A., Wood,W.I. and Godowski,P.
AUTHORS	The Secreted Protein Discovery Initiative (SPDI), a Large-Scale Effort to Identify Novel Human Secreted and Transmembrane Proteins: A Bioinformatics Assessment
JOURNAL	Genome Res. 13 (10), 2265-2270 (2003)
PUBMED	12975109
REFERENCE	2 (bases 1 to 1675) Clark,H.F.
AUTHORS	Direct Submission
TITLE	Submitted (01-AUG-2003) Department of Bioinformatics, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA
JOURNAL	Location/Qualifiers
FEATURES	1..1675 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="DNA88004" 1..1675 /locus_tag="UNQ1934" 337..1173 /locus_tag="UNQ1934" /note="PRO4409" /codon_start=1 /product="CTG4A" /protein_id="AAQ89319.1" /db_xref="GI:37183038" /translation="MDSMPAPSRCLLLPLLLLLLLPAPELGPSQAGABENDWVR LPKRCVKYKVAVELKSAFETGKTEVIGTGYILDOKASGVKTKDRLIETET ICKRLDYSILHKERTGSRNFAKMSSETFTLHNLVHRGVKVMIDPYELWNTEAEVA DLKKOCDLVVEEPEEVEDWYRHOEDLTFLCANHLVKGDTSCLAEOWSGKKGDT ALGKGKSKKSSRAKAGGRSSSKQKELGLEGDPSPDEDSGIQKASPLTHSPPD EL"
ORIGIN	
Query Match	37.6%; Score 691.6; DB 8; Length 1675;
Best Local Similarity	84.5%; Pred. No. 1.9e-158;
Matches 791; Conservative	0; Mismatches 139; Indels 6; Gaps 1;
Qy	5 CGAGGGAGGAGAGCCCGCGGTCCGCTCTGCTCTGGGTCCGGTCCGGCCATGAGTCCA 64
Db	287 CGCGGGAGGAGGAAACCGCCCGGTCCTTAGGGTCCGGGCCCGCGGCCCATGATTCAA 346
Qy	65 TGCTCTGAGCTCGCGCCCGCTGCTCTATTTCCTTTGCTGCTGCTGCTCCGCTGCTGC 124
Db	347 TGCTCTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCCCTTCTGCTGCTGCTGCTGC 406
Qy	125 TCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 184
Db	407 TGCTGCGCGGCCCGGAGCTGGGCCCGAGCCAGGCGGAGCTGAGGAGAACGACTGGGTTT 466
Qy	185 GATTGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTG 244
Db	467 GCCTGCCAGCAAAATGCGAAGTGTGTAATATGTTGCTGTGGAGCTGAAGTCAGCCCTTTG 526
Qy	245 AGGAAACGGGAAAGACCAAGGAAGTGATTGACACCGGCTATGGCATCTCGACCGGAAGG 304
Db	527 AGGAAACCGGCAAGACCAAGAGGTGATTGGCAGCGGCTATGGCATCTTGACCAAGG 586



Db 338 CGCGGAGGAGAAACCGCCGGTCTCTTAGGGTCCGGGCCCGCGGCATGATTCAA 397  
QY 65 TGTCTGAGCTCGCGCCCGCTCTTATTTCTTTCTGCTGCTTCCGCTGCTGC 124  
Db 398 TGCCTGAGCCCGCGTCTTCTGCTTCTTCCCTTCTGCTGCTGCTGCTGC 457  
QY 125 TCCCTTCTGCCCCGAAGCTAGCGCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTGC 184  
Db 458 TGCTCCCGCCCCGAGCTGGGCCCGAGCCAGGCGGAGCTGAGGAGAACGACTGGGTTC 517  
QY 185 GATTGCCAGCAATATCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTG 244  
Db 518 GCCTGCCAGCAATATCGAAGTGTGTAATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG 577  
QY 245 AGGAAACGGGAAGAACCAAGNAATGATTGACACCGGCTATGGCATCTGACCGGAGG 304  
Db 578 AGGAAACGGGAAGAACCAAGGAGGTGATTGGACCGGCTATGGCATCTGACCGAAGG 637  
QY 305 GCTCTGGAGTCAAGTACACCAAGTCCGACTTACGGTTAAATTGAACTCACTGAGACCAATTT 364  
Db 638 CTTCTGGAGTCAATATACCAAGTCCGACTTTCGGTTAATCGAAGTCACTGAGACCAATTT 697  
QY 365 GCAAGAGCTTCTGAGCTACAGCTGCAAGAGGAGAGACTGGCAGCAACCGGTTTGCCA 424  
Db 698 GCAAGAGCTCTCTGATTTATAGCTGCAAGAGGAGAGACCGGAGCAATCGAATTTGCCA 757  
QY 425 AGGATATGTCGAGACCTTTGAGAGCTGCAACACCTAGTCCANAAAGGGTCAAGGTGG 484  
Db 758 AGGCAATGTCAGAGACCTTTGAGACATTTACACACCTGGTACACAAAGGGGTCAAGGTGG 817  
QY 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544  
Db 818 TGATGGACATCCCTATGAGCTGTGGAACGAGACTTCTGAGAGGTGCTGACCTCAAGA 877  
QY 545 AGCAGTGTGAGCTGCTGTGGAAGTTTGAAGAGTGATGAGGACTGGTACAGGAACC 604  
Db 878 AGCAGTGTGATGTGCTGTGGAAGTTTGAAGAGGTGATCGAGGACTGGTACAGGAACC 937  
QY 605 ACCAGGAGGAAGACTGACTGAATTCCTGTGTGCCAACCACTGCTGAAGGGAAGGACA 664  
Db 938 ACCAGGAGGAAGACTGACTGAATTCCTGTGTGCCAACCACTGCTGAAGGGAAGGACA 997  
QY 665 CGAGTTGCTTAGCAGAGCGGTGCTGTGCAAGAGGGGACATAGCTCCCTGGAGGGA 724  
Db 998 CCAAGTTGCTGCGAGAGCAGTGTGTCGCAAGAGGGAGACACAGCTCCCTGGAGGGA 1057  
QY 725 AGAAATCAAGAAGAGCGAGCGAGTCAAG-----GGCTCTTCCAGTGGCAGCAGCA 778  
Db 1058 AGAAGTCCAAGAAGAGCAGCAGGCGCAAGGCGAGCGGCGGAGGATAGCAGCAGCA 1117  
QY 779 AGCAGAGGAAGAACTGGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGGGTG 838  
Db 1118 AACAAAGGAAGAGCTGGGTGGCTTTGAGGAGAGACCCAGGCCCCGAGGAGATGAGGGCA 1177  
QY 839 TGCAGAAGGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGTG 898  
Db 1178 TCCAGAAGGCATCCCTCTCCACACAGCCCCCTGATGAGCTGTGAGCCACCCAGCAT 1237  
QY 899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934  
Db 1238 CTTCTGTCTGAGACCCCTGATTTTGAAGCTGAGGA 1273

RESULT 9  
LOCUS CQ716533 958 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 2467 from Patent WO2068579.  
ACCESSION CQ716533  
VERSION CQ716533.1 GI:42277390  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
1  
Venter, C.J., Adams M.C., Li, P.W. and Myers E.W.  
Kits, such as nucleic acid arrays, comprising a majority of  
humanexons or transcripts, for detecting expression and other uses  
thereof  
Patent: WO 02068579-A 2467 06-SEP-2002;  
PE Corporation (NY) (US)  
FEATURES  
source Location/Qualifiers  
1..958  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 27.3%; Score 502.8; DB 6; Length 958;  
Best Local Similarity 87.7%; Pred. No. 4e-112;  
Matches 549; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 5 CGAGGAGGAAGCGCGCGGTCGGCTCTGCTCTGGGTCCGGCTGGGCCCATGAGTCCA 64  
Db 333 CGCGGAGGAGAAACCGCCCGTCTTTAGGGTCCGGCCCCGGGCCCATGATTCAA 392  
QY 65 TGTCTGAGCTCGCGCCCGCTGCTCTTATTTCTTTCTGCTGCTGCTTCCGCTGCTGC 124  
Db 393 TGCCTGAGCCCGCTCCCGCTGCTCTGCTTCTTCCCTTCTGCTGCTGCTGCTGC 452  
QY 125 TCCCTTCTGCCCCGAAGCTAGGCCCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGC 184  
Db 453 TGTCTCGCGGCCCGGAGCTGGGCCCGAGCCAGCGGAGCTGAGGAGAACGACTGGGTTC 512  
QY 185 GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGGCTTTTG 244  
Db 513 GCCTGCCAGCAATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTTG 572  
QY 245 AGGAAACGGGAAGAACCAAGGAGTGTGACACCGGCTATGGCATCTTGGACCGGAAGG 304  
Db 573 AGGAAACGGGAAGAACCAAGGAGTGTGACACCGGCTATGGCATCTTGGACCGGAAGG 632  
QY 305 GCTCTGAGTCAAGTACACCAAGTCCGACTTACCGTTAAATTGAAAGTCACTGAGACCATTT 364  
Db 633 CTTCTGGAGTCAAAATACACCAAGTCCGACTTTCGGTTAAATCGAAGTCACTGAGACCATTT 692  
QY 365 GCAAGAGCTTCTGAGCTACAGCTGCAAGGAGAGAGCTGGCAGCAACCGGTTTGCCA 424  
Db 693 GCAAGAGCTCTCTGATTTATAGCTTACCAAGGAGAGGACCGGCGAGCAATCGAATTGCCA 752  
QY 425 AGGCTATGTCGGAGACCTTTGAGACGCTGCAACACCTAGTCCACAAAGGGGTCAAGGTGG 484  
Db 753 AGGCAATGTCAGAGACCTTTGAGACATTTACACACCTGGTACACAAAGGGGTCAAGGTGG 812  
QY 485 TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544  
Db 813 TGATGGACATCCCTATGAGCTGTGGAAACGAGACTTCTGCGAGAGGTGCTGACCTCAAGA 872  
QY 545 AGCAGTGTGAGCTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGTGACAGGAACC 604  
Db 873 AGCAGTGTGATGTGCTGGTGAAGAGTTTGAAGAGGTGATCGAGACTGGTACAGGAACC 932  
QY 605 ACCAGGAGGAAGACTGACTGAATTC 630  
Db 933 ACCAGGAGGAAGACTGACTGAATTC 958  
RESULT 10  
LOCUS HSU80744  
DEFINITION Homo sapiens CTG4a mRNA, complete cds.  
ACCESSION U80744  
VERSION U80744.1 GI:2565062  
KEYWORDS  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
1 (bases 1 to 962)  
Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,  
Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.  
cDNAs with long CAG trinucleotide repeats from human brain  
Hum. Genet. 100 (1), 114-122 (1997)  
9225980  
REFERENCE  
AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,  
Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.  
Direct Submission  
Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of  
Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA  
FEATURES  
source Location/Qualifiers  
1..962  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/tissue\_type="cerebral cortex"  
1..962  
/gene="CTG4a"  
388..819  
/gene="CTG4a"  
/note="polyleucine rich"  
/codon\_start=1  
/product="CTG4a"  
/protein\_id="AAB91442.1"  
/db\_xref="GI:2565063"  
/translations="MDSMPESPASRCLLLPLLLLLLLLPAPELGPQSQAAGENDWVR  
LPSKBCVKYAVELASAEETKTEVIGTGILDDQKASGVKYTKSDLRLIETVET  
ICKRLLI IACTRGPAAIDLPRACQRLRHRTTYTKGSRW"  
repeat\_region 439..462  
/rpt\_type=tandem  
/rpt\_unit="ctg"  
ORIGIN  
Query Match 26.7%; Score 491.2; DB 8; Length 962;  
Best Local Similarity 87.3%; Pred. No. 2.8e-109;  
Matches 549; Conservative 0; Mismatches 79; Indels 1; Gaps 1;  
QY 2 GCACGAGGAGGAAGCGCCCGCGGTCTGCTCTGCGTCCGGCTGGGCTGGGCCATGTGAGT 61  
DB 335 GCCCGCGGAGGAGAACCGCCCGTCTTTAGGCTCGGCGCCGCGGCCCATGTGATT 394  
QY 62 CCATGTCTGAGCTCGCGCCCGCTGCTTTATTTCTTTGCTGCTGCTCTCCGCTGC 121  
DB 395 CAATGCTGAGCCGGTCCCGTGTCTTCTGCTTCTTCTCCCTTGTCTGCTGCTGCTGC 454  
QY 122 TGCTCTTCTGCGCGAGCTAGGCCGAGTCCCGCCGGGCTGAGGACACGACTGGG 181  
DB 455 TGCTGTGCGCGCCCGGAGCTGGGCGCGAGCCGAGCTGAGGAACGACTGGG 514  
QY 182 TGGGATTGCCAGCAATGCAAGTGTGCAAGTATGTTGCTGTGGAGCTCAAGTCGGCTT 241  
DB 515 TTGCGCTGCCAGCAATGCGAAGTGTGTAATATATGTTGCTGTGGAGCTGAGTCACT 574  
QY 242 TTGAGGAACCGGAAAGACCAAGGAAGTGTATGACACCGGCTATGGCATCTCGACGGGA 301  
DB 575 TTGAGGAACCGGCAAGACCAAGGAGTGTATGGCACCGGGCTATGGCATCTCGACCCAGA 634  
QY 302 AGGGCTCTGAGTCAAGTACACCAAGTCGACTTACGGTTAATTGAAGTCACTGAGACCA 361  
DB 635 AGGCCTCTGAGTCAAAATACACCAAGTCGACTTGGCGTTAATCGAAGTCACTGAGACCA 694  
QY 362 TTGTCAAGAGGCTTCTGGACTACAGCTGCACAAAGGAGAGGAGTGGCAGCAACCGGTTTG 421  
DB 695 TTGTCAAGAGGCTTCTTCTTATAGCCCTGCACAAAGGAGAGGACCGGACGAAATCGATTG 753  
QY 422 CCAAGGATATGTCGGAGACCTTTTGAGACGCTGCACAACTAGTCTCCACAAAGGGGTCAAGG 481

Db	754	CCAAGGCGCATGTGACAGACCTTTTGACACATTACCAACCTTGATACAAAGGGGTCAAGG	813
Qy	482	TGCTGATGATATCCCTCATGAGCTGTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCA	541
Db	814	TGCTGATGACATCCCTCATGAGCTGTGTGGAACGAGACTTCTGCAGAGGTGGCTGACCTCA	873
Qy	542	AGAGCAGTGTGACGTGCTGGTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGTGACAGGA	601
Db	874	AGAANCAGTGTGTGATGTGCTGGTGTGGAAGAGTTTGAAGAGGTGATCGAGGACTGTGACAGGA	933
Qy	602	ACCACCAGGAGGAAGACCTGACTGAAATTC	630
Db	934	ACCACCAGGAGGAAGACCTGACTGAAATTC	962
RESULT 11			
AC131887			
LOCUS	246404 bp	DNA	linear HTG 09-MAY-2003
DEFINITION	Rattus norvegicus clone CH230-3M14, WORKING DRAFT SEQUENCE, 3		
ACCESSION	AC131887		
VERSION	AC131887.3	GI:30466855	
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 246404)		
	Muzny, D., Marie, E., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayoye, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Cencer, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Geiger, G. E., Geer, K., Gill, R., Grady, M., Guerra, W., Guervara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, J., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Joelivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loshenshew, L., Loutsegh, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puar, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, J., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,		









Db 900 TGAGAAACCGCAAGACCAAGAGGCTGATTGGCAACGGGCTATGGCATCTCGACCAAA 959  
Qy 303 GGGCTCTGGAGTCAAGTACACCAAGTCGAGCTTACGGTTAATTGAAGTCACCTGAGACCAT 362  
Db 960 GGCCTCTGGAGTCAAAATACCAAGTCGAGCTTGGGGTTAATCGAAGTCACCTGAGACCAT 1019  
Qy 363 TTGCAAGAGGCTTCTGGACTACAGCCTCGACCAAGGAGAGGAGCTGGCAGCAACCGGTTTGC 422  
Db 1020 TTGCAAGAGGCTTCTGGATTATAGCTGCAACAAGGAGAGGAGCGCGCAGCAATCGATTGC 1079  
Qy 423 CAAGGGTATGCGGAGACCTTTGAGACGCTGCAACCTAGTCCAACAAAGGGGTCAAGGT 482  
Db 1080 CAAGGGCATGTGAGAGACCTTTGAGACATTACCAACCTGTGTACACAAGGGGTCAAGGT 1139  
Qy 483 GGTGATGATATCCCTATGAGCTGTGGAACGAGACCTTCAGAGAGTGGCTGACCTCAA 542  
Db 1140 GGTGATGATATCCCTATGAGCTGTGGAACGAGACCTTCAGAGAGTGGCTGACCTCAA 1199  
Qy 543 GAAGCAGTGTGACGTGCTGGTGGAAAGCTTTGAAGAGGTGATTGAGGACTGGTACAGGAA 602  
Db 1200 GAAGCAGTGTGATGTGCTGGTGGAAAGCTTTGAGAGAGTGTGAGGAGTGTGAGGACTGGTACAGGAA 1259  
Qy 603 CCACCAGGAGGAGGAGCCTG 621  
Db 1260 CCACCAGGAGGAGGAGCCTG 1278

RESULT 14  
BC008133 LOCUS 1277 bp mRNA linear PRI 28-SEP-2004  
DEFINITION Homo sapiens cDNA clone IMAGE:3457829.  
ACCESSION BC008133  
VERSION BC008133.1 GI:14713265  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1277)  
Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altechul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S., Wang, J., Heide, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1277)  
DIRECTOR MGC Project.  
AUTHORS  
TITLE Direct Submission  
JOURNAL  
Submitted (22-MAY-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.mci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: amgebcm.tmc.edu  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louieged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 19 Row: n Column: 9.

## FEATURES

source

1..1277  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="IMAGE:3457829"  
/tissue\_type="Cervix, carcinoma"  
/clone\_id="NIH\_MGC\_12"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 24.9%; Score 457.8; DB 8; Length 1277;  
Best Local Similarity 81.6%; Pred. No. 4.4e-101; Indels 6; Gaps 1;  
Matches 544; Conservative 0; Mismatches 117;  
Qy 274 GACACCGGCTATGGCATCTCGAGGGAAGGGCTCTGGAGTCAAGTACCAAGTCGGAC 333  
Db 173 GAGCCAGCGCGAGCTGAGGAGACGACTGGTTCGCTGCCAGCAATGCGAAGGGAC 232  
Qy 334 TTACGGTTAATGTAAGTCACTGAGACCATTTGCAAGAGGCTTCTGGACTACAGCTGAC 393  
Db 233 TTGGCGTTAATCGAAGTCACTGAGACCATTTGCAAGAGGCTCCTGGATTATAGCTGAC 292  
Qy 394 AAGGAGGAGGCTGCGACGACCGGTTTCCCAAGGGTATGTCGAGACCTTTGAGACGCTG 453  
Db 293 AAGGAGGAGGCTGCGACGACCATTTGTCGAAGGCGATGTCAGAGACCTTTGAGACATTA 352  
Qy 454 CACAACCTAGTCCCAAGAGGGGTCAAGTGTGTGATGATATCCCTATGAGCTGTGGAAC 513  
Db 353 CACAACCTGTGTACAAAGGGGTCAAGTGTGTGATGACATCCCTATGAGCTGTGGAAC 412  
Qy 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGACGTGCTGGTGGAGAGTTT 573  
Db 413 GAGACTTCTGCAAGGTGGCTGACCTCAAGAGCAGTGTGATGTGCTGGTGGAGAGTTT 472  
Qy 574 GAAGAGGTGATTGAGGCTGTTACAGGAACCAACAGGAGGAGACCTGACCTGAATTCCTC 633  
Db 473 GAGGAGGTGATTGAGGAGCTGGTACAGGAACCAACAGGAGGAGACCTGACCTGAATTCCTC 532  
Qy 634 TGTGCAACACCTGCTGAAGGGAAGGACACGAGTTGCTTAGCAGACGGTGTGCTGGC 693  
Db 533 TGCGCAACACCTGCTGAGGGAAGGACACGAGTTGCTTAGCAGACGGTGTGCTGGC 592  
Qy 694 AAGAGGAGGAGACATAGCCTCCCTGGAGGGAAGAAATCCAAGAGAGAGCGCAGCGAGTC 753  
Db 593 AAGAGGAGGAGACATAGCCTCCCTGGAGGGAAGAAATCCAAGAGAGAGCGCAGCGAGTC 652  
Qy 754 AAG-----GGCTCTCCAGTGGCAGCAGCAGCAGGAGGAGGAGTGGGGGGCTGGG 807  
Db 653 AAGCAGCAGCGCGGAGGAGTAGCAGCAACCAAGAGGAGGAGTGGGTGGCTTGAAG 712  
Qy 808 GAGGATGCCAACCGCGGAGGAGGAGGTGTGCAAGAGGAGGATCGCCCTCCCAACAGC 867  
Db 713 GGAGACCCCGCGGAGGAGGATGAGGGCATCCAGAGGAGGATCCCTCTCTCACACACAGC 772  
Qy 868 CCCCTGTAGTGTGTGAGCCAGCTTGTGTCTTGTGATCAAGACCCCTGACTTCAGAG 927  
Db 773 CCCCTGTAGTGTGTGAGCCAGCTTGTGTCTTGTGATCAAGACCCCTGACTTCAGAG 832

Qy 928 CTTGGGA 934  
 Db 833 CTGAGGA 839

## RESULT 15

BC011767  
 LOCUS Homo sapiens trinucleotide repeat containing 5, mRNA (cDNA clone IMAGE:3609644), partial cds.  
 ACCESSION BC011767  
 VERSION BC011767.2 GI:40226374  
 KEYWORDS

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1593)  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Alekshun, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, W.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932

## TITLE

JOURNAL  
 PUBMED 2 (bases 1 to 1593)  
 Strausberg, R.

## REFERENCE

AUTHORS Direct Submission  
 TITLE Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

## REMARK

COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 On Dec 19, 2003 this sequence version replaced gi:15079962.  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 web site: <http://www.nhsc.nih.gov/>  
 Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brookes, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karling, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 27 Row: f Column: 20

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

## FEATURES

Location/Qualifiers  
 source  
 1..1593  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3609644"  
 /tissue\_type="Uterus, endometrium adenocarcinoma"  
 /clone\_lib="NIH\_MGC\_44"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"

## ORIGIN

Query Match 24.9%; Score 457.8; DB 8; Length 1593;  
 Best Local Similarity 81.6%; Pred. No. 4.4e-101;  
 Matches 544; Conservative 0; Mismatches 117; Indels 6; Gaps 1;  
 Qy 274 GACACCGGCTATGGCATCTCTGACGGGAGGGCTCTGGAGTCAAGTACACCAAGTCGGAC 333  
 Db 475 GAGCCAGGCCGGAGCTGAGGAGAACGACTGGGTTGCGCTGCCAGCAATGCGAAGGGAC 534  
 Qy 334 TTACGGTTAATTGAAGTCACTGAGACCAATTTCGAAGGGCTTCTGGACTACAGCCTGCAC 393  
 Db 535 TTGGCGTTAATCGAAGTCACTGAGACCAATTTCGAAGGGCTTCTGGATTATAGCCTGCAC 594  
 Qy 394 AAGGAGGAGGACTGGCAGCAACCGGTTTTCGAAGGGTATGTCGGAGACCTTTGAGACGCTG 453  
 Db 595 AAGGAGGAGGCGCGCAGCAATCGATTTCGAAGGGCATGTCAGAGACCTTTGAGACATTA 654  
 Qy 454 CACAACCTAGTCCACAAAGGGGTCAAGTGTGTATGATATCCCTATGAGCTGTGGAAC 513  
 Db 655 CACAACCTGTGTACAAAGGGGTCAAGTGTGTATGAGCATCCCTATGAGCTGTGGAAC 714  
 Qy 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGCTGCTGGTGGAGAGTTT 573  
 Db 715 GAGACTTCTGAGAGGTGGCTGACCTCAAGAGCAGTGTGATGTCTGGTGGAGAGTTT 774  
 Qy 574 GAAGAGGTGATTGAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTC 633  
 Db 775 GAGGAGGTGATCGAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTC 834  
 Qy 634 TGTCCCAACCACTGTCTGAAGGGAAGAGACACGAGTTGCTTAGCAGAGCGGTGTCTGGC 693  
 Db 835 TGCGCCAACCACTGTCTGAAGGGAAGAGACACGAGTTGCTTAGCAGAGCAGTGTCTGGC 894  
 Qy 694 AAGAGGGGACATAGCTCCCTCGGAGGGAAGAAATCCAAAGAGAGAGCGCAGCGAGTC 753  
 Db 895 AAGAGGGGAGACACAGCTGCCCTGGGAGGGAAGAGTCCAAGAGAGAGCAGCAGAGGCC 954  
 Qy 754 AAG-----GGCTCTCTCCAGTGGCAGCAGCAGCAGGAGGAAGAACTCGGGGGGCTGGGG 807  
 Db 955 AAGCAGCAGCGCGGAGGAGTAGCAGCAGCAACAAAGAGAGGAGCTGGTGGCTTGGAG 1014  
 Qy 808 GAGAGTCACCAAGCGCGAGGAGGAGGAGGTGTGCAAGAGGAGTATGCCCTCCACACAGC 867  
 Db 1015 GGAGACCCCGAGCCCGGAGGAGTCCAGAGGAGCATCCAGAGGAGTATCCCTCTCACACAGC 1074  
 Qy 868 CCCCTGATGAGCTGTGAGCCCGAGCTTAGTGTCTTGAATCAAGACCCCTGACTTCACAG 927  
 Db 1075 CCCCTGATGAGCTGTGAGCCCGAGCTTAGTGTCTTGAATCAAGACCCCTGACTTCACAG 1134  
 Qy 928 CTTGGGA 934  
 Db 1135 CTGAGGA 1141

Search completed: December 30, 2005, 01:45:58  
 Job time : 9087 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 19:26:35 ; Search time 1051 Seconds  
(without alignments)  
11661.627 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacgaggaggagcgc.....taaaaaaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_21.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001as.\*
- 5: Geneseq2001bs.\*
- 6: Geneseq2002as.\*
- 7: Geneseq2002bs.\*
- 8: Geneseq2003as.\*
- 9: Geneseq2003bs.\*
- 10: Geneseq2003cs.\*
- 11: Geneseq2003ds.\*
- 12: Geneseq2004as.\*
- 13: Geneseq2004bs.\*
- 14: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1839	100.0	1839	4	AAS05278 Murine tr
2	1725.2	93.8	1848	4	AAS05286 DNA seque
3	1528	83.1	1553	14	Adv67441 Nucleotid
4	691.6	37.6	1378	10	ACA54677 Human NF-
5	691.6	37.6	1378	14	ADU83175 Human tri
6	691.6	37.6	1675	8	AAS21311 Human cDN
7	691.6	37.6	1675	8	ACA03670 cDNA enco
8	691.6	37.6	1675	8	ABX89208 DNA encod
9	691.6	37.6	1675	8	ACD41862 Human sec
10	691.6	37.6	1675	8	ACA04091 Human cDN
11	691.6	37.6	1675	9	ADA45654 Novel hum
12	691.6	37.6	1675	9	ADA76085 Human PRO
13	691.6	37.6	1675	9	ADA18735 Human PRO
14	691.6	37.6	1675	9	ADA61358 Homo sapi
15	691.6	37.6	1675	9	ADB19143 Novel hum
16	691.6	37.6	1675	9	ADB27684 cDNA enco
17	691.6	37.6	1675	9	ADA86163 Novel hum
18	691.6	37.6	1675	9	ADB15727 Human PRO
19	691.6	37.6	1675	9	ADA47513 Human PRO

20	691.6	37.6	1675	9	ADA67308 Human PRO
21	691.6	37.6	1675	9	ADB30315 cDNA enco
22	691.6	37.6	1675	9	ADA85611 Novel hum
23	691.6	37.6	1675	9	ADA96823 Human PRO
24	691.6	37.6	1675	9	ADA79127 Human PRO
25	691.6	37.6	1675	9	ADA87266 Novel hum
26	691.6	37.6	1675	9	ADB16468 Human PRO
27	691.6	37.6	1675	9	ADA91560 Novel hum
28	691.6	37.6	1675	9	ADB14623 Human PRO
29	691.6	37.6	1675	9	ADB18584 Novel hum
30	691.6	37.6	1675	9	ADA93799 Human PRO
31	691.6	37.6	1675	9	ADB19695 Novel hum
32	691.6	37.6	1675	9	ADB13007 Human PRO
33	691.6	37.6	1675	9	ACD98491 Novel hum
34	691.6	37.6	1675	9	ADA74261 Human PRO
35	691.6	37.6	1675	9	ADB24494 Human PRO
36	691.6	37.6	1675	9	ADA82018 Human PRO
37	691.6	37.6	1675	9	ADA74981 Human PRO
38	691.6	37.6	1675	9	ADA85059 Novel hum
39	691.6	37.6	1675	9	ADA84507 Novel hum
40	691.6	37.6	1675	9	ADB29763 cDNA enco
41	691.6	37.6	1675	9	ADA80291 Human PRO
42	691.6	37.6	1675	9	ADA75533 Human PRO
43	691.6	37.6	1675	9	ADA46758 Human PRO
44	691.6	37.6	1675	9	ADB25054 Human PRO
45	691.6	37.6	1675	9	ADA93230 Human PRO

#### ALIGNMENTS

RESULT 1  
AAS05278

ID AAS05278 standard; cDNA; 1839 BP.

XX AAS05278;

AC AAS05278;

XX 07-SEP-2001 (first entry)

XX Murine trinucleotide repeat protein (TRP) cDNA sequence.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;

XX transgenic animal; knockout mouse; triplet repeat expansion;

XX fragile X syndrome; Huntington's disease; mouse; murine; ss.

XX Mus sp.

XX Key Location/Qualifiers

FT CDS 55..885

FT /\*tag= a

FT /product= "TRP"

XX WO200130798-A1.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-US029382.

XX 26-OCT-1999; 99US-0161488P.

XX (DELT-) DELTAGEN INC.

XX Klein R, Matthews W, Moore M, Allen KD;

XX WPI; 2001-300473/31.

XX P-PSDB; AAU02498.

XX Novel transgenic animals useful as animal model for characterization of function of a gene encoding trinucleotide repeat proteins (TRPs), contains heterozygous disruption in a gene encoding TRP.

XX Claim 8; Fig 11; 106pp; English.

CC The present sequence encodes for murine trinucleotide repeat protein  
 CC (TRP). The invention describes methods of producing embryonic stem (ES)  
 CC cells comprising a heterozygous disruption in a target DNA sequence  
 CC (preferably gene T243) encoding a TRP and of producing a knockout mouse  
 CC comprising a homozygous disruption in a gene encoding TRP, where the  
 CC disruption inhibits the production of the wild type TRP. The invention  
 CC also relates to identifying agents capable of affecting a phenotype of a  
 CC knockout mouse. Also described are methods of determining whether  
 CC expansion of the trinucleotide repeat in a gene encoding TRP produces a  
 CC phenotypic change. The transgenic animals and the cells are useful for  
 CC identifying compounds capable of ameliorating disease symptoms, and as  
 CC test substrates for the identification of drugs, pharmaceuticals,  
 CC therapies and interventions which may be effective in treating  
 CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's  
 CC disease. The animal models for trinucleotide repeat disorders are ideal  
 CC model systems to study the progression of disease in vivo, the molecular  
 CC basis of these diseases and show the features observed in human disease.  
 CC Using the mice, it is possible to model both the pathogenic mechanism and  
 CC the trinucleotide repeat instability in the mouse

XX  
 Sequence 1839 BP; 424 A; 522 C; 535 G; 358 T; 0 U; 0 Other;

Query Match 100.0%; Score 1839; DB 4; Length 1839;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1839; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGCAGGAGGAGGAGCGCGCGGTCGGTCTGCTCTGGGTCCGGTGGCCATGGAG	60
Db	1	GGCAGGAGGAGGAGCGCGCGGTCGGTCTGCTCTGGGTCCGGTGGCCATGGAG	60
QY	61	TCCATGCTGAGCTCGCGCGCGTCTCTATTTCTTTGCTGCTGCTCTCCGCTG	120
Db	61	TCCATGCTGAGCTCGCGCGCGTCTCTATTTCTTTGCTGCTGCTCTCCGCTG	120
QY	121	CTGCTCTCTCCGCCGAGCTAGGCCGAGTCCCGCGGGCTGAGGACACCGACTGG	180
Db	121	CTGCTCTCTCCGCCGAGCTAGGCCGAGTCCCGCGGGCTGAGGACACCGACTGG	180
QY	181	GTGGATTTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGGAGCTGAAGTCGGCT	240
Db	181	GTGGATTTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGGAGCTGAAGTCGGCT	240
QY	241	TTTGAGGAAAACGGGAAACCAAGAAAGTATGACACCGGTATGGCATCTTGACGGG	300
Db	241	TTTGAGGAAAACGGGAAACCAAGAAAGTATGACACCGGTATGGCATCTTGACGGG	300
QY	301	AAGGCTCTGAGTCAAGTACCAAGTCCGACTTACGGTTAATTGAAGTCACTGAGACC	360
Db	301	AAGGCTCTGAGTCAAGTACCAAGTCCGACTTACGGTTAATTGAAGTCACTGAGACC	360
QY	361	ATTTGCAAGAGGCTTCTGGACTACAGCTGACAGCTGACACCTAGTCCAAAGGGGTCAAG	420
Db	361	ATTTGCAAGAGGCTTCTGGACTACAGCTGACAGCTGACACCTAGTCCAAAGGGGTCAAG	420
QY	421	GCCAAAGGATGTGCGAGACCTTTGAGACGCTGACACCTAGTCCAAAGGGGTCAAG	480
Db	421	GCCAAAGGATGTGCGAGACCTTTGAGACGCTGACACCTAGTCCAAAGGGGTCAAG	480
QY	481	GTGGTGTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTTGGCTGACCTC	540
Db	481	GTGGTGTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTTGGCTGACCTC	540
QY	541	AAGAAGCATGTGACGCTGCTGGTGAAGATTTGAAGAGGTGATTGAGGACTGGTACAGG	600
Db	541	AAGAAGCATGTGACGCTGCTGGTGAAGATTTGAAGAGGTGATTGAGGACTGGTACAGG	600
QY	601	AACACACGAGGAGGAGCTGACTGAAATTCCTCTGTGCAACCAAGTCTGAAGGGAAG	660
Db	601	AACACACGAGGAGGAGCTGACTGAAATTCCTCTGTGCAACCAAGTCTGAAGGGAAG	660
QY	661	GACACGAGTGTGCTAGCAGAGCGGTGGTCTGGCAAGAAAGGGGACATAGCCCTCCCTGGGA	720
Db	661	GACACGAGTGTGCTAGCAGAGCGGTGGTCTGGCAAGAAAGGGGACATAGCCCTCCCTGGGA	720

QY	721	GGGAAAGAAATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAAG	780
Db	721	GGGAAAGAAATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAAG	780
QY	781	CAGAGGAAGAACTGGGGGGCTTGGGGAGGATGCCAAACGCCGAGGAGGAGGAGGTGTG	840
Db	781	CAGAGGAAGAACTGGGGGGCTTGGGGAGGATGCCAAACGCCGAGGAGGAGGAGGTGTG	840
QY	841	CAGAGGCAATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCTGTGAGCCACTTAGTGTG	900
Db	841	CAGAGGCAATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCTGTGAGCCACTTAGTGTG	900
QY	901	CTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACAGCAGCAGCGCAGCGGCGAGCT	960
Db	901	CTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACAGCAGCAGCGCAGCGGCGAGCT	960
QY	961	CCAGCAAGGACAGCTGTGTCAGCATCAGGTCTCTCCCTTGGGTGTGCCCCCTTTCTT	1020
Db	961	CCAGCAAGGACAGCTGTGTCAGCATCAGGTCTCTCCCTTGGGTGTGCCCCCTTTCTT	1020
QY	1021	CCCTTGAACACAGCAAGAGTGGAGGATCTGGGTGCTGGGAGCGCACCCCAAAGG	1080
Db	1021	CCCTTGAACACAGCAAGAGTGGAGGATCTGGGTGCTGGGAGCGCACCCCAAAGG	1080
QY	1081	GAAGAGGAGGAGGAGCAGAGCGAGCTCTCTTCTTACACAGTCCCTCCCTCAGAGCTCCGG	1140
Db	1081	GAAGAGGAGGAGGAGCAGAGCGAGCTCTCTTCTTACACAGTCCCTCCCTCAGAGCTCCGG	1140
QY	1141	GGTCCACCCAGCATCCCGAGCTGAGATCCAGGCTCTGACATGGAAGTGAAGAGCATG	1200
Db	1141	GGTCCACCCAGCATCCCGAGCTGAGATCCAGGCTCTGACATGGAAGTGAAGAGCATG	1200
QY	1201	AGGCACATAGATGCTCACAGCGCCCCCTTCAGCAGGAGAGACTCGGTGCGAGCTCAG	1260
Db	1201	AGGCACATAGATGCTCACAGCGCCCCCTTCAGCAGGAGAGACTCGGTGCGAGCTCAG	1260
QY	1261	CAGCAGGCTGCTCTCTTCCACCAAGCATCTCTTCTGCTGGTCTTGTGCGGATGG	1320
Db	1261	CAGCAGGCTGCTCTCTTCCACCAAGCATCTCTTCTGCTGGTCTTGTGCGGATGG	1320
QY	1321	TAAATTCGAGAACTTCAGGACAAACTCGGGTGTGGCAAAAGGGGTGGACGCGCAGAGC	1380
Db	1321	TAAATTCGAGAACTTCAGGACAAACTCGGGTGTGGCAAAAGGGGTGGACGCGCAGAGC	1380
QY	1381	CAGAGCCACGCGCAGAGCTGCGAGAGGCGCACCTTAACCCCTCCCTGGAAAGCCAAATC	1440
Db	1381	CAGAGCCACGCGCAGAGCTGCGAGAGGCGCACCTTAACCCCTCCCTGGAAAGCCAAATC	1440
QY	1441	TGCAAGTTCCTGTCACCCACTCTCTCTGAGGACGCTCATGTCTTGCCCGAGCCCTTCT	1500
Db	1441	TGCAAGTTCCTGTCACCCACTCTCTCTGAGGACGCTCATGTCTTGCCCGAGCCCTTCT	1500
QY	1501	CCCAGGGCTACAGAGTAAACACACTTTTGGCTTTTGGTGTGGTCTCTGAGTCTCATCA	1560
Db	1501	CCCAGGGCTACAGAGTAAACACACTTTTGGCTTTTGGTGTGGTCTCTGAGTCTCATCA	1560
QY	1561	GCCTCCAGAGTGTCCCTCATCATCTTTTGGCTTTTGGTGTGGTCTCCCGAGGGGCTGG	1620
Db	1561	GCCTCCAGAGTGTCCCTCATCATCTTTTGGCTTTTGGTGTGGTCTCCCGAGGGGCTGG	1620
QY	1621	AAGGCCATCACCATCATTTGGAGGCTTAACCTGTGAGTGTGAGGAGTGTGGAGCGCC	1680
Db	1621	AAGGCCATCACCATCATTTGGAGGCTTAACCTGTGAGTGTGAGGAGTGTGGAGCGCC	1680
QY	1681	CGGGGTTGGTTGGGGTAATCACTCAGTGGCTCTCAGCTTCTAACTGAGCGCCCTTA	1740
Db	1681	CGGGGTTGGTTGGGGTAATCACTCAGTGGCTCTCAGCTTCTAACTGAGCGCCCTTA	1740
QY	1741	ATACAGTTCCTTCTGTTGTGTGTGACTCCACGCGCCCCCACACACACATATAATTT	1800
Db	1741	ATACAGTTCCTTCTGTTGTGTGTGACTCCACGCGCCCCCACACACACATATAATTT	1800







[illegible]

RESULT 4  
ACA54677  
ID ACA5  
XX  
AC ACA5

XX	05-JUN-2003	(first entry)
DT	Human NF-kappaB associated polynucleotide sequence #117.	
XX		
XX	Human; nuclear factor-kappaB; NF-kappaB; immune disorder; cancer;	
KW	inflammatory disorder; apoptosis; hepatic disorder; Hodgkin's lymphoma;	
KW	haematopoietic tumour; hyper-IgM syndrome; viral infection; asthma;	
KW	hypohidrotic ectodermal dysplasia; human immunodeficiency virus; HIV;	
KW	X-linked anhidrotic ectodermal dysplasia; al incontinentia pigmenti;	
KW	influenza; rheumatoid arthritis; inflammatory bowel disease; colitis;	
KW	atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;	
KW	experimental allergic encephalomyelitis; autoimmune disorder; wound;	
KW	hyper immune activity; acute phase response; hypercongenital condition;	
KW	birth defect; necrotic lesion; organ transplant rejection; pancreas;	
KW	signal transduction; hyperproliferative disorder; diabetes mellitus;	
KW	vitamin B12 malabsorption; neurological disorder; Huntington's chorea;	
KW	Turner's syndrome; bacterial infection; cardiovascular disorder;	
KW	infertility; psoriasis; haemolytic anaemia; antiinflammatory; anti-HIV;	
KW	cytostatic; hepatotropic; virucide; antirheumatic; antiarthritic;	
KW	antidiabetic; immunomodulator; antidiabetic; antiallergic;	
KW	neuroprotective; immunosuppressive; vulnery; antibacterial;	
KW	antiinfectivity; antitanaemic; antipsoriatic; cerebroprotective; cardiac;	
KW	antiarteriosclerotic; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
WO	WO200286076-A2.	
XX		
PD	31-OCT-2002.	
XX		
PF	19-APR-2002; 2002WO-US012636.	
XX		
PR	19-APR-2001; 2001US-0284962P.	
XX		
PR	26-APR-2001; 2001US-0286645P.	
XX		
PR	09-JAN-2002; 2002US-0346986P.	
XX		
PA	(BRIM ) BRISTOL-MYERS SQUIBB CO.	
XX		
PI	Carman J, Feder J, Nadler S;	
XX		
DR	WPI; 2003-0931119/08.	
XX		
DR	P-PSDB; ABU69616.	
XX		
PT	Novel NF-kappaB-associated polypeptides and polynucleotides useful for	
PT	diagnosing, treating and preventing cancer, hepatic disorders, aberrant	
PT	apoptosis, viral infections, autoimmune disorders, asthma and stroke.	
XX		
PS	Claim 4; Fig 10H; 608pp; English.	
XX		
CC	The present invention relates to the isolation of human nuclear factor-	
CC	kappaB (NF-kappaB) associated polypeptides and polynucleotides. The NF-	
CC	kappaB associated polypeptide and polynucleotide sequences are useful for	
CC	preventing, treating or ameliorating various disorders including immune	
CC	disorders, inflammatory disorders, cancers, disorders relating to	
CC	aberrant apoptosis, hepatic disorders, Hodgkin's lymphomas,	
CC	haematopoietic tumours, hyper-IgM syndromes, hypohidrotic ectodermal	
CC	dysplasia, X-linked anhidrotic ectodermal dysplasia, immunodeficiency, al	
CC	incontinentia pigmenti, viral infections (e.g. those caused by human	
CC	immunodeficiency virus (HIV), human T-cell lymphotropic virus (HTLV),	
CC	hepatitis B, hepatitis C, Epstein Barr virus (EBV), influenza),	
CC	rheumatoid arthritis, inflammatory bowel disease, colitis, asthma,	
CC	atherosclerosis, cachexia, euthyroid sick syndrome, stroke, experimental	
CC	allergic encephalomyelitis (EAE), autoimmune disorders, disorders related	
CC	to hyper immune activity, disorders related to aberrant acute phase	
CC	responses, hypercongenital conditions, birth defects, necrotic lesions,	
CC	wounds, organ transplant rejection, disorders related to aberrant signal	
CC	transduction, hyperproliferative disorders, diseases of the pancreas	
CC	(e.g. diabetes mellitus, vitamin B12 malabsorption), neurological	
CC	disorders (e.g. Huntington's chorea), Turner's syndrome, bacterial	
CC	infections, cardiovascular disorders, infertility, psoriasis and	
CC	haemolytic anaemia. The present sequence represents a human NF-kappaB	
CC	associated polynucleotide of the invention	

XX	SQ	Sequence	1378 BP; 311 A; 399 C; 419 G; 249 T; 0 U; 0 Other;
		Query Match	37.6%; Score 691.6; DB 10; Length 1378;
		Best Local Similarity	84.5%; Pred. No. 6.4e-168;
		Matches	791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;
Qy	5	CGAGGAGGAAGCCGCCGGTTCGTCTGCTCTGGGTCCGGCTCAGACCAATGAGTGCCA	64
Db	21	CGCGGAGGAGAAACCGCCCGTCTTTAGGGTCCGGGCCCGCGGCATCATGATTCAA	80
Qy	65	TGCTGAGCTCGCGCCCGCTGCTCTTAATTTCTTTGCTGCTGCTCTCGCTGCTGTC	124
Db	81	TGCTGAGCCGCGTCCGCTGCTCTTCTCTCTTCCCCTTCTGCTGCTGCTGCTGCTGC	140
Qy	125	TCCTTCTTCGCCCGAAGCTAGGCCCGAGTCCCGCCGGGCTTAGAGACCACTGGGTGC	184
Db	141	TGCTGCCGCCCGGAGCTGGCCCGAGCCAGCCGGAGCTGAGGAGAACGACTGGGTTC	200
Qy	185	GATTGCCAGCAATATGCAAGTGTGCAAGTATATGTTGCTGTGGAGCTGAAFTCGGCTTTG	244
Db	201	GCCTGCCAGCAATATGCAAGTGTATAATATGTTGCTGTGGAGCTGAAFTCAGCCTTTG	260
Qy	245	AGGAAACGGGAAGACCAAGAAAGTATGACACGGCTATGGCATCTCTGACCGGAGG	304
Db	261	AGGAAAACGGGAAGACCAAGAGAGTATTTGGCACCGGCTATGGCATCTCTGACCAGAAG	320
Qy	305	GCTCTGAGTCAAGTATACCAAGTCGGAATTCAGGTAAATTTGAAGTCACTGAGACCAATTT	364
Db	321	CCTCTGAGTCAATATACCAAGTCGGAATTCAGGTAAATTCGAAGTCACTGAGACCAATTT	380
Qy	365	GCAAGAGCTTCTGGACTPACAGCTGCAACAGCTGCAACAGGAGAGCACTGGCAGCAACCGTTTGGCA	424
Db	381	GCAAGAGCTCCTGGATATATAGCTTGCAAGAGCTGCAACACCTAGTTCACAAGGGGTCAAGTGG	440
Qy	425	AGGCTATGTCGGAGACCTTTGAGAGCTGCAACACCTAGTTCACAAGGGGTCAAGTGG	484
Db	441	AGGCAATGTCAGAGACCTTTGAGACATTTACACAACCTGGTATACACAAGGGGTCAAGTGG	500
Qy	485	TGATGGATATCCCCCTATGAGCTGTGGAAACGAGACTCAGCAGAGGTGGCTACCTCAAGA	544
Db	501	TGATGGACATCCCCCTATGAGCTGTGGAAACGAGACTTCTGACAGAGTGGCTGACCTCAAGA	560
Qy	545	AGCAGTGTAGCTGCTGGTGAAGAGTTTGAAGAGGTGATTTGAGGACTGTGTTACAGGAACC	604
Db	561	AGCAGTGTATGTGCTGTGAAGAGTTTGAAGAGGTGATTCAGGAGCTGTGTTACAGGAACC	620
Qy	605	ACGAGGAGGAGACTGACTCAAAATTCCTCTGTGCCAACCACTGCTGAAAGGAAAGGACA	664
Db	621	ACGAGGAGGAGACTGACTCAAAATTCCTCTGTGCCAACCACTGCTGAAAGGAAAGGACA	680
Qy	665	CGAGTGTCTAGCAGAGCGGTGGTCTGGCAAGAAAGGGGACATAGCTTCCTCGGAGGGGA	724
Db	681	CCAGTGTCTGGCAGAGCAGTGGTTCGGCAAGAAAGGGGAGACACAGCTGCCCTGGAGGGGA	740
Qy	725	AGAAATCCAGAAGACGCGCAGCGAGTCAAG-----GGTCTCTCCAGTGGCAGCAGCA	778
Db	741	AGAAATCCAGAAGACGCGCAGCGGCAAGGCAAGGCAAGCGGCGGAGGATAGCAGCAGCA	800
Qy	779	AGCAGAGGAGGAACCTGGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGGGTG	838
Db	801	AACAAAAGGAGGAGCTGGTGGCTCTGAGGGAGAGCCCCCAGCCCCGAGAGGATGAGGGA	860
Qy	839	TGCAGAAGGCATCGCCCTCCCAACACAGCCCCCTGTATGAGCTGTGAGCCCCAGCTTAGTG	898
Db	861	TCCAGAAGGCATCCCTCTCACACACGCCCCCTGTATGAGCTGTGAGCCCCACCCAGCAT	920
Qy	899	TCCTTGAATCAAGACCCCTGACTTCAAGACTTTGGGA	934
Db	921	CCTCTGTCTTGAGACCCCTGATTTTGAAGCTTGAGGA	956

## RESULT 5

ADU831175  
ID ADU831175 standard; cDNA; 1378 BP.  
XX  
XX AC ADU831175;  
XX  
XX DT 10-FEB-2005 (first entry)  
XX  
XX DE Human trinucleotide repeat containing 5 protein, cDNA.  
XX  
XX KW Nuclear factor kappa B; NFkappaB; inflammation; cardiovascular disease; neoplasm; gastrointestinal disease; immune disorder; immune deficiency; XX  
XX KW dermatological disease; infection; nutritional disorder;  
XX KW cerebrovascular ischemia; endocrine disease; injury; respiratory disease;  
XX KW gynecology and obstetrics; kawasaki disease; rheumatic fever; diagnosis; ss; gene.  
XX  
XX OS Homo sapiens.  
XX  
XX FN WO2004100886-A2.  
XX  
XX PD 25-NOV-2004.  
XX  
XX PF 06-MAY-2004; 2004WO-US014279.  
XX  
XX PR 07-MAY-2003; 2003US-00431096.  
XX  
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
XX PI Carman J, Feder JN, Nadler SG;  
XX  
XX DR WPI; 2005-047909/05.  
XX  
XX DR P-PSDB; ADU831186.  
XX  
XX PT New nucleic acid molecule encoding an NFkappaB associated molecule, useful for diagnosing, preventing, treating, or ameliorating a medical condition, e.g. cancer, wounds, or immune, inflammatory, hepatic, viral or pulmonary disorder.  
XX  
XX PS Claim 1; SEQ ID NO 139; 646pp; English.  
XX  
XX CC The invention relates to an isolated nucleic acid molecule having NFkappaB (nuclear factor kappa B) modulating activity (or its fragments, variants, homologs and sequences hybridizing to it). Also included are an isolated polypeptide capable of modulating an NFkappaB response (or its domain, epitope, variant, species homolog or interacting protein), an isolated antibody that binds specifically to the isolated polypeptide, a method for preventing (treating, or ameliorating) a medical condition, a method of diagnosing a NFkappaB associated condition (or a susceptibility to a NFkappaB associated condition) in a subject, a method for identifying a binding partner to the polypeptide, a method of identifying a compound that modulates the biological activity of a NFkappaB associated molecule, a method of screening for a compound that is capable of modulating the biological activity of a NFkappaB associated molecule and a compound that modulates the biological activity of a human NFkappaB associated molecule as identified by any of the methods above. The nucleic acid molecule and polypeptides, composition and methods are useful for diagnosing, preventing, treating, or ameliorating a medical condition, e.g. immune disorder, an inflammatory disorder in which polypeptides of the present invention are associated with the disorder either directly or indirectly, an inflammatory disorder related to aberrant NFkappaB regulation, a cancer, aberrant apoptosis, hepatic disorders, Hodgkins lymphomas, hematopoietic tumors, hyper-IGM syndromes, hypohydrotic ectodermal dysplasia, X-linked anhidrotic ectodermal dysplasia, immunodeficiency, al incontinentia pigmenti, viral infections, HIV-1, HTLV-1, hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell survival, and evasion of immune responses, rheumatoid arthritis, inflammatory bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper immune activity, disorders related to aberrant acute phase responses, hypercongenital conditions, birth defects, necrotic lesions, wounds, organ transplant rejection, conditions related to organ transplant rejection, disorders related to aberrant signal transduction, proliferating disorders, HIV

CC propagation in cells infected with other viruses, associated with EL-8,  
 CC disorders associated with aberrant IL-8 expression, disorders associated  
 CC with aberrant IL-8 activity, pulmonary disorders, pulmonary fibrosis,  
 CC Behcet's disease, bacterial infections, gynecological diseases,  
 CC psoriasis, IgA nephropathy, chronic obstructive pulmonary disease,  
 CC Kawasaki disease, Crohn's disease, peripheral arterial occlusive disease,  
 CC Hodgkin's disease, idiopathic intermediate uveitis, hyaline membrane  
 CC disease, acute rheumatic fever, chronic rheumatic heart disease,  
 CC ulcerative colitis, autoimmune disorders, and autoimmune thyroid disease.  
 CC The present sequence encodes a protein found to interact with the NFkappa  
 CC B modulating protein AD037.

XX Sequence 1378 BP; 311 A; 399 C; 419 G; 249 T; 0 U; 0 Other;

Query Match 37.6%; Score 691.6; DB 14; Length 1378;  
 Best Local Similarity 84.5%; Pred. No. 6.4e-168;  
 Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

QY	5	CGAGGAGGAGGAGCGCGCGGCTCGCTCTGCTGGGTTCGGCTGGGCGCATGGAGTCCA	64
DB	21	CGCGGAGGAGGAGAACCGCGCGTCTTAGGGTCCGGCGCGCGCGCCATGGATTCAA	80
QY	65	TGCTGAGCTCGCGCGCGCGCTGCTTATTTTCTGCTGCTGCTGCTGCTGCTGCTG	124
DB	81	TGCCTGAGCGCGCGCGCTGCTGCTTCTTCCCTTGTGCTGCTGCTGCTGCTGCTG	140
QY	125	TCCTTCCTCCCGGAGCTAGCGCGGCTAGCGCGGCTGAGGAGACCGACTGGGTGC	184
DB	141	TCCTCGCGCGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	200
QY	185	GATTGCCAGCAATATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGGCTTTG	244
DB	201	GCTTCCCGCAGCAATATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTTG	260
QY	245	AGGAACCGGAGGAGGAGGAGGAGTGTGACACCGCTGATGGCATCTGACCGGAGG	304
DB	261	AGGAACCGGAGGAGGAGGAGGAGTGTGACACCGCTGATGGCATCTGACCGGAGG	320
QY	305	GCTCTGAGTCAAGTACACCAAGTCCGACTTACGGTTAATGAACTCACTGAGACATTT	364
DB	321	CCTCTGAGTCAATATACCAAGTCCGACTTACGGTTAATGAACTCACTGAGACATTT	380
QY	365	GCAAGAGCTTCTGAGCTACAGCTGCAAGGAGGAGTGTGACAGCAACCGGTTTGCA	424
DB	381	GCAAGAGCTTCTGAGCTACAGCTGCAAGGAGGAGTGTGACAGCAACCGGTTTGCA	440
QY	425	AGGATGCTCGGAGACCTTGGAGCGCTGACAACTAGTCCACAAAGGGGTCAAGTGG	484
DB	441	AGGATGCTCGGAGACCTTGGAGCGCTGACAACTAGTCCACAAAGGGGTCAAGTGG	500
QY	485	TGATGGATATCCCTATGAGCTGTGGACGAGACCTCAGCAGAGGTGCTGACCTCAAGA	544
DB	501	TGATGGACATCCCTATGAGCTGTGGACGAGACCTCAGCAGAGGTGCTGACCTCAAGA	560
QY	545	AGCAGTGTGAGCTGTGTTGAGAGTGTGAGAGGTGATGAGGACTGGTACAGGAACC	604
DB	561	AGCAGTGTGAGCTGTGTTGAGAGTGTGAGAGGTGATGAGGACTGGTACAGGAACC	620
QY	605	ACCAGGAGGAGACCTGACTGATTTCTTCTGTCGCAACACAGTGTGAGGAGGAGACA	664
DB	621	ACCAGGAGGAGACCTGACTGATTTCTTCTGTCGCAACACAGTGTGAGGAGGAGACA	680
QY	665	CGAGTTGCTAGCAGAGCGGTGCTGTCGCAAGAGGGGACATAGCTTCCCTGGGAGGGA	724
DB	681	CGAGTTGCTAGCAGAGCGGTGCTGTCGCAAGAGGGGACATAGCTTCCCTGGGAGGGA	740
QY	725	AGAAATCCAAAGAGGAGCGAGCGAGTCAAG-----GGCTCTCTCCAGTGGCAGCACA	778
DB	741	AGAAATCCAAAGAGGAGCGAGCGAGTCAAG-----GGCTCTCTCCAGTGGCAGCACA	800
QY	779	AGCAGAGGAGGAAATGGGGGGCTTGGGGGGAGGATGCCAAAGCCCGAGGAGGAGGAGTG	838
DB	801	AACAAGGAGGAGCTGGGTGGCTTTGAGGGAGACCCCGAGCGCGCGGAGGATGAGGGCA	860

QY	839	TGCAGAGGAGATCCGCCCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGTG	898
DB	861	TCCAGAGGAGATCCGCCCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCCAGCAT	920
QY	899	TCCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGA	934
DB	921	CCTCTGCTCTGAGACCCCTGATTTTGAAGCTGAGGA	956

RESULT 6

AAS21311

ID AAS21311 standard; cDNA; 1675 BP.

AC AAS21311;

XX 24-OCT-2001 (first entry)

XX Human cDNA sequence encoding for PRO4409 polypeptide.

DE Human secretory and transmembrane; PRO; mammalian; cancer; lung; breast;  
 KW prostate; cervical; tumour necrosis factor-alpha; TNF-alpha; cartilage;  
 KW eat; proliferation; glucose; free fatty acid; skeletal muscle; adipocyte;  
 KW A-peptide; factor VIIA; gene therapy; ss.

XX Homo sapiens.

OS WO200140466-A2.

PN 07-JUN-2001.

PD 01-DEC-2000; 2000WO-US032678.

PR 01-DEC-1999; 99WO-US028301.

PR 01-DEC-1999; 99WO-US028634.

PR 02-DEC-1999; 99WO-US028551.

PR 02-DEC-1999; 99WO-US028564.

PR 02-DEC-1999; 99WO-US028565.

PR 09-DEC-1999; 99US-0170262P.

PR 16-DEC-1999; 99WO-US030095.

PR 20-DEC-1999; 99WO-US030911.

PR 20-DEC-1999; 99WO-US030999.

PR 30-DEC-1999; 99WO-US031243.

PR 30-DEC-1999; 99WO-US031274.

PR 06-JAN-2000; 2000WO-US000219.

PR 06-JAN-2000; 2000WO-US000277.

PR 11-FEB-2000; 2000WO-US000376.

PR 18-FEB-2000; 2000WO-US004341.

PR 18-FEB-2000; 2000WO-US004342.

PR 24-FEB-2000; 2000WO-US004914.

PR 24-FEB-2000; 2000WO-US005004.

PR 01-MAR-2000; 2000WO-US005601.

PR 02-MAR-2000; 2000WO-US005841.

PR 03-MAR-2000; 2000US-0187202P.

PR 10-MAR-2000; 2000WO-US006319.

PR 15-MAR-2000; 2000WO-US006884.

PR 20-MAR-2000; 2000WO-US007377.

PR 21-MAR-2000; 2000WO-US007532.

PR 30-MAR-2000; 2000WO-US008439.

PR 17-MAY-2000; 2000WO-US013705.

PR 30-MAY-2000; 2000WO-US014941.

PR 02-JUN-2000; 2000WO-US015364.

PR 05-JUN-2000; 2000US-0209832P.

PR 28-JUL-2000; 2000WO-US020710.

PR 11-AUG-2000; 2000WO-US022031.

PR 23-AUG-2000; 2000WO-US023522.

PR 24-AUG-2000; 2000WO-US023328.

PR 08-NOV-2000; 2000WO-US030952.

PR 10-NOV-2000; 2000WO-US030873.

XX

PA	(GETH ) GENENTECH INC.	
XX	Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;	
PI	Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;	
PI	Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;	
XX	WPI; 2001-408281/43.	
DR	P-PSDB; AAU12239.	
XX	Isolated, secretory and transmembrane PRO polypeptide used to detect	
PT	other PRO polypeptides, link bioactive molecules to cells expressing PRO	
PT	polypeptides, and detect the presence of mammalian tumors e.g. lung,	
PT	breast, prostate, cervical.	
XX	Claim 3; Fig 135; 813pp; English.	
PS	AA521244-AAS21518 encode for novel human secretory and transmembrane PRO	
CC	polypeptides. The PRO polypeptides are useful to detect other PRO	
CC	polypeptides, to link bioactive molecules to cells expressing PRO	
CC	polypeptides, to modulate biological activities of cells expressing PRO	
CC	polypeptides, and to detect the presence of mammalian lung, colon,	
CC	breast, prostate, rectal, cervical or liver tumours by comparing PRO	
CC	polypeptide expression in a cell sample to that in a control sample. Some	
CC	of the 275 sequences are also useful to stimulate the release of tumour	
CC	necrosis factor-alpha (TNF-alpha) from human blood, the proliferation or	
CC	differentiation of chondrocytes, the proliferation or gene expression in	
CC	pericyte cells, the release of proteoglycans from cartilage, the	
CC	proliferation of inner ear utricular supporting cells or of T-	
CC	lymphocytes, the release of a cytokine from peripheral blood monocytes	
CC	(PMBCs), or the proliferation of endothelial cells. Some of the PRO	
CC	polypeptides may modulate glucose or free fatty acid uptake by skeletal	
CC	muscle cells or by adipocytes; or inhibit binding of A-peptide to factor	
CC	VIIA. The PRO polypeptides can be used in assays to identify molecules	
CC	involved in binding interactions. The polynucleotides encoding PRO	
CC	polypeptides can be used to generate probes, antisense RNA/DNA,	
CC	transgenic or knock out animals and can be used in gene therapy	
XX	Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 U; 0 Other;	
SQ		
QY	Query Match 37.6%; Score 691.6; DB 4; Length 1675;	
DB	Best Local Similarity 84.5%; Pred. No. 7e-168;	
QY	Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;	
DB	5 CGAGGGAGGAGCGCGCGGTCGCTCTGCTGGGTCCGGCTGGGCCATGAGTCCA 64	
QY	287 CGCGGAGGAGAACCGCCGCTCTTAGGTCGCGGCCGCGGCCATGATTCAA 346	
DB	65 TGTCTGAGTCGCGCGCGCTCTTATTTCTTTCTGCTGCTGCTGCTGCTGCTG 124	
QY	347 TGCCTGAGCGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 406	
DB	125 TCCTTCTGCGCGAGCTAGCCCGAGTCGCCGCGCGGCTGAGAGACCGACTGGGTGC 184	
QY	407 TGTCTGCGCGCGCGGCTGCGCGCGCGCGCGCGAGCTGAGGAGAACGACTGGTTC 466	
DB	185 GATTGCCAGCAATGCCAGTGTGCAAGTATGTTGCTGAGCTGAAGTCGGCTTTTG 244	
QY	467 GCCTGCCAGCAATGCCAGTGTGCAAGTATGTTGCTGAGCTGAAGTCAGCCTTTG 526	
DB	245 AGGAAACGGGAAAGAACCAAGAGTGTGACACCGGCTATGGCATCTGGACGGGAAGG 304	
QY	527 AGGAAACGGGAAAGAACCAAGAGTGTGACACCGGCTATGGCATCTGGACGGGAAGG 586	
DB	305 GCTCTGAGTCAAGTATACCAAGTCGGACTTACCGTTAATTGAAGTCACTGAGACCAATT 364	
QY	587 CTTCTGAGTCAAGTATACCAAGTCGGACTTACCGTTAATTGAAGTCACTGAGACCAATT 646	
DB	365 GCAGAGGCTTCTGCACTACGCTGCAAGGAGGACTGGCAGCAACCGTTTGCCA 424	
QY	647 GCAAGAGGCTTCTGCACTACGCTGCAAGGAGGACTGGCAGCAACCGTTTGCCA 706	
DB	425 AGGGTATGTCGAGACCTTTGAGACGCTGCAACCTAGTTCACAAAGGGGTCAAGGTGG 484	
QY		

Db 707 AGGCATGTGTCAGAGACCTTTGAGACATTAACAACCTGGGTACACAAGGGGTCAAGGTGG 766

QY 485 TGATGGATATCCCTTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGA 544

Db 767 TGATGGATATCCCTTATGAGCTGTGGAAACGAGACTTCTGACAGAGTGGCTGACCTCAAGA 826

QY 545 AGCAGTGTGAGCTGCTGGTGGAGAGTGTGAAGAGGTGATTGAGGACTGTGTTACAGGAACC 604

Db 827 AGCAGTGTGAGTGTGCTGGTGGAGAGTGTGAGGAGTGTGATCGAGGACTGTGTTACAGGAACC 886

QY 605 ACCAGGAGGAGAGACTGACTGAATTCCTGTGTGCCAACCAAGCTGCTGAGGAGGAGGAGCA 664

Db 887 ACCAGGAGGAGAGACTGACTGAATTCCTGTGTGCCAACCAAGCTGCTGAGGAGGAGGAGCA 946

QY 665 CGAGTTCCTAGCAGAGCGGTGTGCTGCGCAAGAGGGGGACATAGCTCCTCCCTGGGAGGGA 724

Db 947 CCAGTTCCTGGCAGAGCAGTGTGCTCGGCAAGAGGAGGAGACACAGCTGCCCTGGGAGGGA 1006

QY 725 AGAATCCAGAGAGAGCGCAGCGGAGTCAAG-----GGCTCTCTCCAGTGGCAGCAGCA 778

Db 1007 AGAAGTCCAGAGAGAGCAGCAGGCGGCAAGGCGCAGCGGCGGAGGAGTAGCAGCAGCA 1066

QY 779 AGCAGAGAGAGAACTGGGGGCTGGGGGAGGATGCCAACCGCAGGAGGAGGAGGAGGTG 838

Db 1067 AACAAAGGAGAGGAGCTGGGTGGCTTTGAGGGAGACCCCGCAGGAGGATGAGGGCA 1126

QY 839 TGCAAGAGGATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCCGAGCTTAGTGTG 898

Db 1127 TCCAGAGGATCCCTCTCTACACAGCCCCCTGATGAGCTGTGAGCCCGAGCTTAGTGTG 1186

QY 899 TCCCTGAATCAAGACCCCTGACTTTCAGAGCTTGGGA 934

Db 1187 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 1222

RESULT 7

ACA03670

ID ACA03670 standard; cDNA; 1675 BP.

XX ACA03670;

AC ACA03670;

XX 23-MAY-2003 (first entry)

DT

XX cDNA encoding human PRO polypeptide #68.

DE Human; PRO polypeptide; secreted and transmembrane protein;

KW tumour necrosis factor-alpha; TNF-alpha; blood; proliferation;

KW differentiation; chondrocyte; tumour; genetic disorder; cytostatic; gene;

XX ss.

XX Homo sapiens.

OS

XX US2003036180-A1.

PN

XX 20-FEB-2003.

PD

XX 09-MAY-2002; 2002US-00143114.

PF

XX 31-MAR-1997; 97WO-US005230.

PR 12-JUN-1998; 98WO-US012456.

PR 14-JUL-1998; 98WO-US014552.

PR 28-AUG-1998; 98WO-US017888.

PR 10-SEP-1998; 98WO-US018824.

PR 14-SEP-1998; 98WO-US019093.

PR 14-SEP-1998; 98WO-US019094.

PR 16-SEP-1998; 98WO-US019330.

PR 17-SEP-1998; 98WO-US019437.

PR 07-OCT-1998; 98WO-US021141.

PR 29-OCT-1998; 98WO-US022591.

PR 20-NOV-1998; 98WO-US024855.

PR 01-DEC-1998; 98WO-US025108.





01-MAR-2001; 2001WO-US006666.  
09-MAR-2001; 2001US-00802706.  
14-MAR-2001; 2001US-00808689.  
22-MAR-2001; 2001US-00816744.  
05-APR-2001; 2001US-00828366.  
10-MAY-2001; 2001US-00854208.  
10-MAY-2001; 2001US-00854280.  
18-MAY-2001; 2001US-00860216.  
25-MAY-2001; 2001US-00866028.  
25-MAY-2001; 2001US-00866034.  
25-MAY-2001; 2001WO-US017092.  
01-JUN-2001; 2001US-00872035.  
01-JUN-2001; 2001WO-US017800.  
05-JUN-2001; 2001US-00874503.  
14-JUN-2001; 2001US-00882636.  
19-JUN-2001; 2001US-00886342.  
20-JUN-2001; 2001WO-US019692.  
21-JUN-2001; 2001US-00887879.  
22-JUN-2001; 2001WO-US020116.  
29-JUN-2001; 2001WO-US021066.  
09-JUL-2001; 2001WO-US021735.  
18-JUL-2001; 2001US-00908827.  
06-AUG-2001; 2001US-00924419.  
09-AUG-2001; 2001US-00927796.  
16-AUG-2001; 2001US-00931836.  
19-DEC-2001; 2001US-00028072.  
XX  
XX  
(GENTH ) GENENTECH INC.  
XX  
XX  
Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;  
P1 Grittens ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
P1 Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
XX  
XX WPI; 2003-148238/14.  
DR P-PSDB; ABU59718.  
XX  
XX Two hundred and seventy five nucleic acids encoding PRO polypeptides,  
PT useful for treating pericyte-associated tumors, diabetes and various bone  
PT and/or cartilage disorders, e.g. arthritis.  
XX  
XX Claim 2; Fig 135; 659pp; English.  
XX  
XX The invention describes an isolated human PRO polypeptide. The PRO  
CC polypeptides are useful in detecting PRO polypeptides in a sample, in  
CC linking a bioactive molecule to a cell expressing a PRO polypeptide, and  
CC in modulating at least one biological activity of a cell expressing a PRO  
CC polypeptide. PRO1312 stimulates hypertrophy of neonatal heart and is thus  
CC useful for treating cardiac insufficiency disorders. PRO1154 and PRO1186  
CC stimulate adrenal cortical capillary endothelial growth, and PRO536,  
CC PRO943, PRO828, PRO826, PRO1068 or PRO535, PRO826, PRO819, PRO1126,  
CC PRO1360 and PRO1387 induce c-fos in endothelial cells, and are thus  
CC useful for treating conditions or disorders where angiogenesis would be  
CC beneficial, e.g. wound healing and antagonist of this polypeptide are  
CC useful for treating cancerous tumors. PRO812 inhibits vascular  
CC endothelial growth factor (VEGF) stimulated proliferation of endothelial  
CC cells and is thus useful for inhibiting endothelial cell growth in  
CC mammals which would be beneficial in inhibiting tumour growth. PRO826,  
CC PRO1068, PRO1184, PRO1346 and PRO1375 stimulate proliferation of  
CC stimulated T-lymphocytes and are therapeutically useful for enhancing  
CC immune response. PRO828, PRO1068 or PRO1132 enhance survival of  
CC retinal neurons cells (PRO1132 is also enhances survival/proliferation of  
CC rod photoreceptor cells) and therefore are useful for treating retinal  
CC disorders of injuries, e.g. retinitis pigmentosa, AMD. PRO819, PRO813  
CC and PRO11066 induce proliferation of mammalian kidney mesangial cells,  
CC and therefore are useful for treating kidney disorders associated with  
CC decreased mesangial cell function such as Berger disease or other  
CC nephropathies associated with diabetes, herpeticiformis or Crohn's  
CC disease. PRO1310, PRO844, PRO1312, PRO1192 and PRO1387 induce the  
CC proliferation and/or redifferentiation of chondrocytes in culture and are  
CC thus useful for treating sports injuries, and arthritis. This sequence  
CC encodes a novel human PRO protein  
XX  
XX Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 U; 0 Other;  
SQ

XX ACD41862;  
AC 05-SEP-2003 (first entry)  
XX Human secreted/transmembrane protein (PRO) cDNA #68.  
XX  
XX Human; ss; Gene; PRO; secreted protein; transmembrane protein; tumour;  
KW cytosolic; gene therapy; tumour necrosis factor-alpha; TNF-alpha; blood;  
KW proteoglycan; cartilage; cytokine; peripheral blood mononuclear cell;  
KW PBMC; glucose uptake; rFA; skeletal muscle cell; adipocyte cell;  
KW chondrocyte cell proliferation; chondrocyte cell differentiation;  
KW pericyte cell; inner ear utricular supporting cell; T-lymphocyte cell;  
KW endothelial cell; A-peptide; factor VIIA.  
XX  
OS Homo sapiens.  
XX  
XX US2003036179-A1.  
XX  
XX 20-FEB-2003.  
XX  
XX 10-MAY-2002; 2002US-00142431.  
XX  
XX 31-MAR-1997; 97WO-US005230.  
PR 12-JUN-1998; 98WO-US012456.  
PR 14-JUL-1998; 98WO-US014552.  
PR 28-AUG-1998; 98WO-US017888.  
PR 10-SEP-1998; 98WO-US018824.  
PR 14-SEP-1998; 98WO-US019093.  
PR 14-SEP-1998; 98WO-US019094.  
PR 14-SEP-1998; 98WO-US019177.  
PR 16-SEP-1998; 98WO-US019330.  
PR 17-SEP-1998; 98WO-US019437.  
PR 07-OCT-1998; 98WO-US021141.  
PR 29-OCT-1998; 98WO-US022991.  
PR 29-OCT-1998; 98WO-US022992.  
PR 20-NOV-1998; 98WO-US024855.  
PR 01-DEC-1998; 98WO-US025108.  
PR 05-JAN-1999; 98WO-US000106.  
PR 08-MAR-1999; 99WO-US005028.  
PR 10-MAR-1999; 99WO-US005190.  
PR 20-APR-1999; 99WO-US008615.  
PR 14-MAY-1999; 99WO-US010733.  
PR 02-JUN-1999; 99WO-US012252.  
PR 01-SEP-1999; 99WO-US020111.  
PR 08-SEP-1999; 99WO-US020594.  
PR 13-SEP-1999; 99WO-US020944.  
PR 15-SEP-1999; 99WO-US021090.  
PR 15-SEP-1999; 99WO-US021547.  
PR 05-OCT-1999; 99WO-US023089.  
PR 29-NOV-1999; 99WO-US028214.  
PR 30-NOV-1999; 99WO-US028313.  
PR 30-NOV-1999; 99WO-US028409.  
PR 01-DEC-1999; 99WO-US028301.  
PR 01-DEC-1999; 99WO-US028634.  
PR 02-DEC-1999; 99WO-US028551.  
PR 02-DEC-1999; 99WO-US028564.  
PR 02-DEC-1999; 99WO-US028565.  
PR 16-DEC-1999; 99WO-US030095.  
PR 20-DEC-1999; 99WO-US030911.  
PR 20-DEC-1999; 99WO-US030999.  
PR 22-DEC-1999; 99WO-US030720.  
PR 30-DEC-1999; 99WO-US031243.  
PR 30-DEC-1999; 99WO-US031274.  
PR 05-JAN-2000; 2000WO-US000219.  
PR 06-JAN-2000; 2000WO-US000277.  
PR 06-JAN-2000; 2000WO-US000376.  
PR 11-FEB-2000; 2000WO-US003565.  
PR 18-FEB-2000; 2000WO-US004341.  
PR 18-FEB-2000; 2000WO-US004342.  
PR 22-FEB-2000; 2000WO-US004414.  
PR 24-FEB-2000; 2000WO-US004914.  
PR 24-FEB-2000; 2000WO-US005004.  
PR 01-MAR-2000; 2000WO-US005601.  
PR 02-MAR-2000; 2000WO-US005746.  
PR 10-MAR-2000; 2000WO-US005841.  
PR 15-MAR-2000; 2000WO-US006319.  
PR 20-MAR-2000; 2000WO-US006884.  
PR 21-MAR-2000; 2000WO-US007377.  
PR 21-MAR-2000; 2000WO-US007532.  
PR 30-MAR-2000; 2000WO-US008439.  
PR 17-MAY-2000; 2000WO-US013705.  
PR 22-MAY-2000; 2000WO-US014042.  
PR 30-MAY-2000; 2000WO-US014941.  
PR 02-JUN-2000; 2000WO-US015264.  
PR 28-JUL-2000; 2000WO-US020710.  
PR 11-AUG-2000; 2000WO-US022031.  
PR 23-AUG-2000; 2000WO-US023522.  
PR 24-AUG-2000; 2000WO-US023328.  
PR 08-NOV-2000; 2000WO-US030952.  
PR 10-NOV-2000; 2000WO-US030873.  
PR 01-DEC-2000; 2000WO-US032678.  
PR 20-DEC-2000; 2000US-00747259.  
PR 20-DEC-2000; 2000WO-US034956.  
PR 28-FEB-2001; 2001US-00796498.  
PR 28-FEB-2001; 2001WO-US006520.  
PR 01-MAR-2001; 2001WO-US006666.  
PR 09-MAR-2001; 2001US-00802706.  
PR 14-MAR-2001; 2001US-00808689.  
PR 22-MAR-2001; 2001US-00816744.  
PR 05-APR-2001; 2001US-00828366.  
PR 10-MAY-2001; 2001US-00854208.  
PR 10-MAY-2001; 2001US-00854280.  
PR 18-MAY-2001; 2001US-00860216.  
PR 25-MAY-2001; 2001US-00866028.  
PR 25-MAY-2001; 2001US-00866034.  
PR 25-MAY-2001; 2001WO-US017092.  
PR 01-JUN-2001; 2001US-00872035.  
PR 01-JUN-2001; 2001WO-US017800.  
PR 05-JUN-2001; 2001US-00874503.  
PR 14-JUN-2001; 2001US-00882636.  
PR 19-JUN-2001; 2001US-00886342.  
PR 20-JUN-2001; 2001WO-US019692.  
PR 21-JUN-2001; 2001US-00887879.  
PR 22-JUN-2001; 2001WO-US020116.  
PR 29-JUN-2001; 2001WO-US021066.  
PR 09-JUL-2001; 2001WO-US021735.  
PR 18-JUL-2001; 2001US-00908827.  
PR 06-AUG-2001; 2001US-00924419.  
PR 09-AUG-2001; 2001US-00927796.  
PR 16-AUG-2001; 2001US-00931836.  
PR 19-DEC-2001; 2001US-00028072.  
XX (GETH ) GENENTECH INC.  
XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;  
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
XX  
XX WPI; 2003-466355/44.  
XX P-PSDB; ABO24908.  
XX  
XX New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO1114 or  
PT PR04978, useful in molecular biology, chromosome and gene mapping, in  
XX generating antisense RNA and DNA, and in gene therapy.  
XX  
XX Claim 2; Fig 135; 659pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising at least 80%  
CC sequence identity to a PRO (secreted and transmembrane protein) cDNA  
CC comprising a nucleic acid (a) encoding a PRO polypeptide, or its  
CC extracellular domain (with or without its associated signal peptide),  
CC which comprises any of the 275 120-850 residue amino acid sequences,  
CC given in the specification; (b) comprising any of the 275 300-3500  
CC nucleotide sequences, given in the specification; or (c) comprising the  
CC full-length coding sequence of the nucleotide sequences given in the







```
Db 467 GCCTCCCGAGCAATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTGG 526
Qy 245 AGGAAACGGGAAAGCAACGAAGTGAATGACACCGGCTATGSCATCTCTGACGGGAAGG 304
Db 527 AGGAAACCGGCAAGCAACGAAGGTGATTTGGCACGGCTATGGCATCTCTGACCAAGAAG 586
Qy 305 GCTCTGGAGTCAAGTACACCAAGTCGAGCTTAAGGTTAATTGAAGTCACTGAGACCATTT 364
Db 587 CCTCTGGAGTCAAAATACACCAAGTCGAGCTTTCGGGTTAATCGAAGTCACTGAGACCATTT 646
Qy 365 CCAAGAGCTCTCTGAGCTACAGCTGCAACGAGGAGGACTGGCAGCAACCGTTTGCCA 424
Db 647 GCAAGAGCTCTCTGATTATAGCTTGCACAGGAGGAGGACCGCAGCAATCGATTTGCCA 706
Qy 425 AGGATATGTCGAGAGCACTTTGAGAGCGCTGCACAACTAGTCCACAAAGGGGTCAGGTTGG 484
Db 707 AGGCAATGTCAGAGACCTTTGAGACATTTACACAACTGTTACACAAAGGGTCAAGTTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGAGCGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGAGCGAGACTTCTGCAGAGGTGCTGACCTCAAGA 826
Qy 545 AGCAGTGTGAGTGTGCTGTGGAAGTGTGAGAGGTGATTCAGGACTGTTACAGGAC 604
Db 827 AGCAGTGTGATGTGCTGTGGAAGAGTTTGAGAGGTGATTCAGGACTGTTACAGGAC 886
Qy 605 ACCAGGAGAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAAGGGAAGGACA 664
Db 887 ACCAGGAGAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAAGGGAAGGACA 946
Qy 665 CGAGTTGCTAGCAGAGCGGTGCTGCGAAGAGGGGACATAGCCTCCCTGGAGGGA 724
Db 947 CCAGTTGCTGCGCAGAGCGGTGCTGCGAAGAGGGGACACAGCTGCCCTGGAGGGA 1006
Qy 725 AGAATCCAAGAGAGCGCAGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778
Db 1007 AGAAGTCCAAGAGAGAGCAGCAGGCGCAAGGCGGCGGAGGAGTACAGCAGCA 1066
Qy 779 AGCAGAGAGAGAACTGGGGGCGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGGGTG 838
Db 1067 AACAAAGAGAGAGTGGGTGGCTTGGAGAGACCCAGCCCGGAGAGATGAGGGA 1126
Qy 839 TGCAGAGGCAATGCCCTCCACACAGCCCGCTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db 1127 TCCAGAGGCAATCCCTCTCACACAGCCCGCTGATGAGCTGTGAGCCCGAGCAGCAT 1186
Qy 899 TCCTTGATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 1187 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 1222

RESULT 11
ADA45654
ID ADA45654 standard; cDNA; 1675 BP.
XX
AC ADA45654;
XX
DT 20-NOV-2003 (first entry)
XX
DE Novel human secreted and transmembrane protein PRO4409 cDNA.
XX
KW Human; secreted and transmembrane protein; PRO; gene; ss;
KW Tumour necrosis factor alpha release; TNF-alpha release;
KW glucose uptake modulator; FFA uptake modulator;
KW cell proliferation stimulator; cell differentiation stimulator;
KW cell differentiation inhibitor; cytokine release stimulator; tumour;
KW lung tumour; colon tumour; breast tumour; prostate tumour; rectal tumour;
KW cervical tumour; liver tumour; chromosome mapping; gene mapping;
KW gene therapy; chromosome identification; chromosome marker.
XX
OS Homo sapiens.
XX
```

```
PN US2003022328-A1.
XX 30-JAN-2003.
XX 16-APR-2002; 2002US-00123904.
XX 31-MAR-1997; 97WO-US005230.
PR 12-JUN-1998; 98WO-US012456.
PR 14-JUL-1998; 98WO-US014552.
PR 28-AUG-1998; 98WO-US017888.
PR 10-SEP-1998; 98WO-US018824.
PR 14-SEP-1998; 98WO-US019093.
PR 14-SEP-1998; 98WO-US019094.
PR 14-SEP-1998; 98WO-US019177.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98WO-US019437.
PR 07-OCT-1998; 98WO-US021141.
PR 29-OCT-1998; 98WO-US022991.
PR 29-OCT-1998; 98WO-US022992.
PR 20-NOV-1998; 98WO-US024855.
PR 01-DEC-1998; 98WO-US025108.
PR 05-JAN-1999; 99WO-US000106.
PR 08-MAR-1999; 99WO-US005028.
PR 10-MAR-1999; 99WO-US005190.
PR 20-APR-1999; 99WO-US008615.
PR 14-MAY-1999; 99WO-US010733.
PR 02-JUN-1999; 99WO-US012252.
PR 01-SEP-1999; 99WO-US020111.
PR 08-SEP-1999; 99WO-US020594.
PR 13-SEP-1999; 99WO-US020944.
PR 15-SEP-1999; 99WO-US021090.
PR 15-SEP-1999; 99WO-US021547.
PR 05-OCT-1999; 99WO-US023089.
PR 29-NOV-1999; 99WO-US028214.
PR 30-NOV-1999; 99WO-US028313.
PR 30-NOV-1999; 99WO-US028409.
PR 01-DEC-1999; 99WO-US028501.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028651.
PR 02-DEC-1999; 99WO-US028564.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 20-DEC-1999; 99WO-US030999.
PR 22-DEC-1999; 99WO-US030720.
PR 30-DEC-1999; 99WO-US031243.
PR 30-DEC-1999; 99WO-US031274.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000277.
PR 06-JAN-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 18-FEB-2000; 2000WO-US004342.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US005004.
PR 01-MAR-2000; 2000WO-US005601.
PR 02-MAR-2000; 2000WO-US005746.
PR 02-MAR-2000; 2000WO-US005841.
PR 10-MAR-2000; 2000WO-US006319.
PR 15-MAR-2000; 2000WO-US006884.
PR 20-MAR-2000; 2000WO-US007377.
PR 21-MAR-2000; 2000WO-US007532.
PR 30-MAR-2000; 2000WO-US008439.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 28-JUL-2000; 2000WO-US020710.
PR 11-AUG-2000; 2000WO-US022031.
PR 23-AUG-2000; 2000WO-US023522.
PR 24-AUG-2000; 2000WO-US023328.
PR 08-NOV-2000; 2000WO-US030952.
```

CC	affinity purification of PRO from recombinant cell culture or natural
CC	sources. (I) and (II) are useful for tissue typing. This sequence encodes
CC	a novel human secreted and transmembrane PRO polypeptide.
XX	
SQ	Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 U; 0 Other;
	Query Match 37.6%; Score 691.6; DB 9; Length 1675;
	Best Local Similarity 84.5%; Pred. No. 7e-168;
	Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;
Qy	5 CGAGGGAGGAAGCCGCCCGGTCCTGCTCTGGTTCGGTTCGGGCGATGAGTCCA 64
Db	287 CGCGGAGGAGGAACCGCCCGGTCCTTTAGGGTCCGGCGCGGCGCATGGAATCAA 346
Qy	65 TGTCTGAGCTCGCGCCCGGTCCTCTTATTTCTTTGCTGCTCTGCTCGCTGCTGC 124
Db	347 TGCCTGAGCCCGGTCCTGCTGCTCTTCTGCTTCTCCCTTGTCTGCTGCTGCTGCT 406
Qy	125 TCCTTCTGCTCCCGGAAGCTAGGCCCGGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGC 184
Db	407 TGTCTGCGCGCCCGGAGCTGGCCCGAGCCAGCCGAGCTGAGGAGAACGACTGGGTTC 466
Qy	185 GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTG 244
Db	467 GCCTGCCAGCAATGCGAAGTGTGTAATATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG 526
Qy	245 AGGAACGGGAAGAACCAAGGAAGTGAATGACACCGGCTATGGCATCTGGACCGGAAGG 304
Db	527 AGGAACCGGCAAGCAAGAGAGTGAATGCGCGGCTATGGCATCTGGACCGGAAGG 586
Qy	305 GCTCTGGAGTCAAGTACACCAAGTCGGACTTTACGGTTAAATTTGAAGTCACTGAGACCAATTT 364
Db	587 CCTCTGAGTCAATATACCAAGTCGGACTTTGCGTTAAATCGAAGTCACTGAGACCAATTT 646
Qy	365 GCAAGAGCTTCTGGACTACAGCTGCAACAAGAGAGAGACTGGCAGCAACCGGTTTGCCA 424
Db	647 GCAAGAGCTTCTGGATTATAGCTTACAGCTGCAACAAGAGAGAGCCGGCAGCAATCGATTTGCCA 706
Qy	425 AGGTATGTCGGAGACCTTTGAGAGCTGCAACACCTAGTCCACAAGGGGTCAAGGTGG 484
Db	707 AGGCATGTGAGAGACCTTTGAGACATTTACACAACCTGGTATCAACAAGGGGTCAAGGTGG 766
Qy	485 TGATGATATCCCTATGAGCTGTGGAAACGAGACCTTCAGCAGAGGTGGCTGACCTCAAGA 544
Db	767 TGATGACATCCCTATGAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA 826
Qy	545 AGCAGTGTGAGCTGCTGGTGGAGAGATTGGAAGGTGATTTGAGAGCTGTGTACAGAACCC 604
Db	827 AGCAGTGTGATGTCTGTTGGAAGAGTTTGAAGAGGTGATCGAGGAGTGGGTACAGGAACC 886
Qy	605 ACCAGGAGGAGACCTGACTGAATTCCTGTGTCCAAACCTGCTGAAGGGAAGGACA 664
Db	887 ACCAGGAGGAGACCTGACTGAATTCCTGTGTCCAAACCTGCTGAAGGGAAGGACA 946
Qy	665 CGAGTGTGCTAGCAGAGCGGTGGTCTGGCAAGAAGGGGACATAGCCTTCCCTGGGAGGGA 724
Db	947 CCAGTGTGCTGGCAGAGCAGTGGTCCGCAAGAAGGGGAGACACAGTCCCTGGGAGGGA 1006
Qy	725 AGAAATCAAGAAGAACCGCAGCGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778
Db	1007 AGAAGTCCAAGAAGAGCAGCAGCGGCGCAAGGCGGAGGAGGAGTATGAGCAGCAGCA 1066
Qy	779 AGCAGAGGAGGAGTGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGAGGGGTG 838
Db	1067 AACAAAGGAGGAGTGGGTGGCTTGAAGGAGACCCCGCCCGAGGAGGAGTATGAGGGA 1126
Qy	839 TGCAGAAGGAGTCCGCTCCACACAGACCCCTCTGATGAGCTGTGAGGCCAGCTTAGTG 898
Db	1127 TCCAGAGGAGTCCCTCTTCAACACAGACCCCTCTGATGAGCTCTGAGGCCACCCAGCAT 1186
Qy	899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db	1187 CCTCTGCTGAGAGCCCTCTGATTTTGAAGCTGAGGA 1222

PR 10-NOV-2000; 2000WO-US030873.

PR 01-DEC-2000; 2000WO-US032678.

PR 20-DEC-2000; 2000US-00747259.

PR 20-DEC-2000; 2000WO-US034956.

PR 28-FEB-2001; 2001US-00796498.

PR 28-FEB-2001; 2001WO-US006520.

PR 01-MAR-2001; 2001WO-US006666.

PR 09-MAR-2001; 2001US-00802706.

PR 14-MAR-2001; 2001US-00808689.

PR 22-MAR-2001; 2001US-00816744.

PR 05-APR-2001; 2001US-00828366.

PR 10-MAY-2001; 2001US-00854208.

PR 10-MAY-2001; 2001US-00854280.

PR 18-MAY-2001; 2001US-00860216.

PR 25-MAY-2001; 2001US-00866028.

PR 25-MAY-2001; 2001US-00866034.

PR 25-MAY-2001; 2001WO-US017092.

PR 01-JUN-2001; 2001US-00872035.

PR 01-JUN-2001; 2001WO-US017800.

PR 05-JUN-2001; 2001US-00874503.

PR 14-JUN-2001; 2001US-00882636.

PR 19-JUN-2001; 2001US-00886342.

PR 20-JUN-2001; 2001WO-US019692.

PR 21-JUN-2001; 2001US-00887879.

PR 22-JUN-2001; 2001WO-US020116.

PR 29-JUN-2001; 2001WO-US021066.

PR 09-JUL-2001; 2001WO-US021735.

PR 18-JUL-2001; 2001US-00908827.

PR 06-AUG-2001; 2001US-00924419.

PR 09-AUG-2001; 2001US-00927796.

PR 16-AUG-2001; 2001US-00931836.

PR 19-DEC-2001; 2001US-00028072.

PA (GETH ) GENENTECH INC.

XX

XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;

PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;

PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

XX

DR WPI; 2003-584997/55.

DR P-PSDB; ADA45655.

XX

XX Novel secreted and transmembrane polypeptide for modulating biological

PT activity of cell expressing the polypeptide, identifying agonists or

PT antagonists of polypeptide, and as molecular weight markers.

XX

PS Claim 2; Fig 135; 659pp; English.

XX

CC The invention describes 305 nucleic acids encoding PRO (secreted and

CC transmembrane) polypeptides (I). (I) is useful for stimulating the

CC release of TNF-alpha from human blood, for modulating the uptake of

CC glucose or FFA by skeletal muscle cells or adipocyte cells, for

CC stimulating the proliferation or differentiation of chondrocyte cells,

CC for stimulating the proliferation of or gene expression in pericyte

CC cells, for stimulating the release of proteoglycans from cartilage, for

CC stimulating the proliferation of inner ear utricular supporting cells,

CC for stimulating the proliferation of T-lymphocyte cells, for stimulating

CC the release of a cytokine from PBMC cells, for inhibiting the binding of

CC A-peptide to factor VIIA, for inhibiting the differentiation of adipocyte

CC cells, for stimulating proliferation of endothelial cells, for detecting

CC the presence of tumour in a mammal. The tumour is lung, colon, breast,

CC prostate, rectal, cervical or liver tumour. The oligonucleotide probes

CC are useful for isolating genomic and cDNA nucleotide sequences or

CC antisense probes. (I) is also useful as therapeutic agent. PRO is useful

CC in assays to identify other proteins or molecules involved in binding

CC interaction. A polynucleotide (II) encoding (I) is useful in chromosome

CC and gene mapping, in generation of antisense RNA and DNA, in the

CC preparation of PRO polypeptide, for generating transgenic animals or

CC knockout animals which in turn are useful in the development and

CC screening of therapeutically useful reagents, in gene therapy, for

CC chromosome identification, as chromosome marker, and for generating

CC probes. An anti-(I)-antibody is useful in diagnostic assays for PRO, e.g.

CC detecting its expression in specific cells, tissues or serum, and for

## RESULT 12

ADA76085

ID ADA76085 standard; cDNA; 1675 BP.

XX AC ADA76085;

XX DT 20-NOV-2003 (first entry)

XX DE Human PRO polynucleotide #68.

XX KW Human; gene; ss; PRO; secreted polypeptide; transmembrane polypeptide;

KW tumour necrosis factor-alpha; TNF-alpha; chondrocyte cell; tumour;

KW cancer; adrenal; lung; colon; breast; prostate; rectum; kidney; cervix;

KW liver; microvascular endothelial cell; glucose; FFA;

KW skeletal muscle cell; adipocyte cell; pericyte cell;

KW inner ear utricular supporting cell; T-lymphocyte cell;

KW endothelial cell tube formation; bone disorder; cartilage disorder;

KW sports injury; proteoglycan; articular cartilage defect; osteoarthritis;

KW rheumatoid arthritis; haemoglobin-associated disorder thalassaemia;

KW immune system cell infiltration.

XX OS Homo sapiens.

XX PN US2003073212-A1.

XX PD 17-APR-2003.

XX PF 16-APR-2002; 2002US-00123903.

XX PR 31-MAR-1997; 97WO-US005230.

XX PR 12-JUN-1998; 98WO-US012456.

XX PR 14-JUL-1998; 98WO-US014552.

XX PR 28-AUG-1998; 98WO-US017888.

XX PR 10-SEP-1998; 98WO-US018824.

XX PR 14-SEP-1998; 98WO-US019093.

XX PR 14-SEP-1998; 98WO-US019094.

XX PR 14-SEP-1998; 98WO-US019177.

XX PR 16-SEP-1998; 98WO-US019330.

XX PR 17-SEP-1998; 98WO-US019437.

XX PR 07-OCT-1998; 98WO-US021141.

XX PR 29-OCT-1998; 98WO-US022991.

XX PR 29-OCT-1998; 98WO-US022992.

XX PR 20-NOV-1998; 98WO-US024855.

XX PR 01-DEC-1998; 98WO-US025108.

XX PR 05-JAN-1999; 99WO-US000106.

XX PR 08-MAR-1999; 99WO-US005028.

XX PR 10-MAR-1999; 99WO-US005190.

XX PR 20-APR-1999; 99WO-US008615.

XX PR 14-MAY-1999; 99WO-US010733.

XX PR 02-JUN-1999; 99WO-US012252.

XX PR 01-SEP-1999; 99WO-US020111.

XX PR 08-SEP-1999; 99WO-US020594.

XX PR 13-SEP-1999; 99WO-US020944.

XX PR 15-SEP-1999; 99WO-US021090.

XX PR 15-SEP-1999; 99WO-US021547.

XX PR 05-OCT-1999; 99WO-US023089.

XX PR 29-NOV-1999; 99WO-US028214.

XX PR 30-NOV-1999; 99WO-US028313.

XX PR 30-NOV-1999; 99WO-US028409.

XX PR 01-DEC-1999; 99WO-US028301.

XX PR 01-DEC-1999; 99WO-US028634.

XX PR 02-DEC-1999; 99WO-US028551.

XX PR 02-DEC-1999; 99WO-US028564.

XX PR 02-DEC-1999; 99WO-US028565.

XX PR 16-DEC-1999; 99WO-US030095.

XX PR 20-DEC-1999; 99WO-US030399.

XX PR 22-DEC-1999; 99WO-US030720.

XX PR 30-DEC-1999; 99WO-US031243.

XX PR 30-DEC-1999; 99WO-US031274.

XX PR 05-JAN-2000; 2000WO-US000219.

PR 06-JAN-2000; 2000WO-US000277.  
 PR 06-JAN-2000; 2000WO-US000376.  
 PR 11-FEB-2000; 2000WO-US003565.  
 PR 18-FEB-2000; 2000WO-US004341.  
 PR 18-FEB-2000; 2000WO-US004342.  
 PR 22-FEB-2000; 2000WO-US004414.  
 PR 24-FEB-2000; 2000WO-US004914.  
 PR 24-FEB-2000; 2000WO-US005004.  
 PR 01-MAR-2000; 2000WO-US005601.  
 PR 02-MAR-2000; 2000WO-US005746.  
 PR 02-MAR-2000; 2000WO-US005841.  
 PR 10-MAR-2000; 2000WO-US006319.  
 PR 15-MAR-2000; 2000WO-US006884.  
 PR 20-MAR-2000; 2000WO-US007377.  
 PR 21-MAR-2000; 2000WO-US007532.  
 PR 30-MAR-2000; 2000WO-US008439.  
 PR 17-MAY-2000; 2000WO-US013705.  
 PR 22-MAY-2000; 2000WO-US014042.  
 PR 30-MAY-2000; 2000WO-US014941.  
 PR 02-JUN-2000; 2000WO-US015264.  
 PR 28-JUL-2000; 2000WO-US020710.  
 PR 11-AUG-2000; 2000WO-US022031.  
 PR 23-AUG-2000; 2000WO-US023522.  
 PR 24-AUG-2000; 2000WO-US023328.  
 PR 08-NOV-2000; 2000WO-US030952.  
 PR 10-NOV-2000; 2000WO-US030873.  
 PR 01-DEC-2000; 2000WO-US032678.  
 PR 20-DEC-2000; 2000US-00747259.  
 PR 20-DEC-2000; 2000WO-US034956.  
 PR 28-FEB-2001; 2001US-00796498.  
 PR 28-FEB-2001; 2001WO-US006520.  
 PR 01-MAR-2001; 2001WO-US006666.  
 PR 09-MAR-2001; 2001US-00802706.  
 PR 14-MAR-2001; 2001US-00808689.  
 PR 22-MAR-2001; 2001US-00816744.  
 PR 05-APR-2001; 2001US-00828366.  
 PR 10-MAY-2001; 2001US-00854208.  
 PR 18-MAY-2001; 2001US-00860216.  
 PR 25-MAY-2001; 2001US-00866028.  
 PR 25-MAY-2001; 2001US-00866034.  
 PR 01-JUN-2001; 2001US-00872035.  
 PR 01-JUN-2001; 2001WO-US017800.  
 PR 05-JUN-2001; 2001US-00874503.  
 PR 14-JUN-2001; 2001US-00882636.  
 PR 19-JUN-2001; 2001US-00886342.  
 PR 20-JUN-2001; 2001WO-US019892.  
 PR 21-JUN-2001; 2001US-00887879.  
 PR 22-JUN-2001; 2001WO-US020116.  
 PR 29-JUN-2001; 2001WO-US021066.  
 PR 09-JUL-2001; 2001WO-US021735.  
 PR 18-JUL-2001; 2001US-00908827.  
 PR 06-AUG-2001; 2001US-00924419.  
 PR 09-AUG-2001; 2001US-00927796.  
 PR 16-AUG-2001; 2001US-00931836.  
 PR 19-DEC-2001; 2001US-00028072.  
 XX (GETH ) GENENTECH INC.  
 PA Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;  
 XX Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
 XX WPI; 2003-687639/65.  
 DR P-PSDB; ADA76086.  
 XX New isolated nucleic acid encoding a secreted and transmembrane  
 PT polypeptide, designated e.g. PRO1114 or PRO4978, useful in chromosome and  
 PT gene mapping, in generating antisense RNA and DNA, and in gene therapy.  
 XX Claim 2; Fig 135; 659pp; English.  
 PS  
 XX

The invention relates to isolated human PRO polypeptides (secreted and transmembrane polypeptides) and the polynucleotides encoding them. The invention also relates to an antibody which specifically binds to a PRO polypeptide, a method for stimulating the release of tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) from human blood, a method for stimulating the proliferation or differentiation of chondrocyte cells and a method for detecting the presence of a tumour in a mammal (e.g. adrenal, lung, colon, breast, prostate, rectal, kidney, cervical and liver tumours). The polynucleotides are useful in molecular biology, including uses as hybridisation probes, in chromosome and gene mapping, in generating antisense RNA and DNA and in gene therapy. The polynucleotides may also be used in preparing PRO polypeptides by recombinant techniques and in generating either transgenic animals or knock-out animals which are useful in the development and screening of therapeutically useful reagents. The PRO polypeptides or antibodies are used in preparing a medicament for treating a condition responsive to the polypeptides or antibodies, such as tumours, for stimulating and inhibiting proliferation of human microvascular endothelial cells, for modulating the uptake of glucose or FFA by skeletal muscle cells or adipocyte cells, for stimulating differentiation of adipocyte cells, for stimulating proliferation or gene expression in pericyte cells, for stimulating the proliferation of inner ear uricular supporting cells or T-lymphocyte cells, for inducing endothelial cell tube formation and for treating various bone and/or cartilage disorders such as sports injuries and arthritis. PRO polypeptides which stimulate the release of proteoglycans from cartilage are useful for treating sports-related joint problems, articular cartilage defects, osteoarthritis and rheumatoid arthritis. PRO polypeptides are also useful for treating various mammalian haemoglobin-associated disorders such as various thalasaemias and conditions which may benefit from enhanced local immune system cell infiltration. This sequence represents a human PRO polynucleotide of the invention. Note: The sequence data for this patent is also available in electronic format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).

Sequence 1675 BP: 380 A: 469 C: 520 G: 306 T: 0 U: 0 Other: 0

Query Match	37.6%	Score 691.6	DB 9	Length 1675	
Best local Similarity	84.5%	Prod. No. 7e-168			
Matches	791	Conservative 0	Mismatches 139	Indels 6	Gaps 1
QY	5	CGAGGGAGGAAGCGCGCGGGTCCGCTCTCTCTGCGTCCGGTGGGCCCATCGAGTCCA	64		
DB	287	CGCGGGAGGAGAACCGCCCGGTCTTTAGGGTCGGGCCCGCGGCCCATGATTCAA	346		
QY	65	TGCTCTGAGCTCGCGCCCGCTGSCCTCTTAATTTCCCTTCTGCTCTGCTCTCCGCTGCTGC	124		
DB	347	TGCTCTGAGCCCGCTCCCGCTCTCTCTGCTCTTCCCTTCTGCTCTGCTGCTGCTGC	406		
QY	125	TCCTTCTGTCGCCAAGCTAGGCCCGAGTCCCGCCGGGGCTGAGGAGACCGACTGGGTGC	184		
DB	407	TGCTGCCGCCCGCGAGCTGGGCCCGAGCCAGGCCGGAGCTGAGGAGAACGACTGGGTTC	466		
QY	185	GATTGCCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGCTGGAGCTGAAGTCGCGCTTTG	244		
DB	467	GCCTGCCCAGCAATGCGAAGTGTGTAATATGTTGCTGTGGAGCTGAAGTCAGCGCTTTG	526		
QY	245	AGGAAACGGGAAGAACCAAGAGTGATTGACACCGGCTATTGGCATCTCGACCGGAAGG	304		
DB	527	AGGAAACGGCAAGAACCAAGAGGTGATTGGCACCGGGCTATTGGCATCTCGACCGAAGG	586		
QY	305	GCTCTGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATTGAAGTCACTGAGACCAATT	364		
DB	587	CCTCTGGAGTCAATACACCAAGTCGGACTTGGGTTAATCGAAGTCACTGAGACCAATT	646		
QY	365	GCAAGAGGCTTCTGACTTACAGCTTGCAACAAGGAGAGGACTGGGCAAGAACCGGTTTGCA	424		
DB	647	GCAAGAGGCTCTGGATTATAGCCTTGCAACAAGGAGAGGACCGGCAGCAATCGATTGCGCA	706		
QY	425	AGGGTATGTCGAGACCTTTGAGAGCTTGCAACCTAGTCCACAAAGGGGTCAAGTGG	484		
DB	707	AGGGCATGTCAGAGACCTTTGAGACATTTACACACCTGGTATACAAAGGGGTCAAGGTGG	766		
QY	485	TGATGGATATCCCTTATGAGCTGTGGAACCGAGACCTCAGCAGAGAGTGGCTGACCTCAAGA	544		









Db 1127 TCAGAGGATCCCTCTCACACAGCCCCCTGATGAGCTGTGAGCCACCAGCAT 1186

Qy 899 TCCTTGATCAAGACCCCTGACTTCAGAGCTTGGGA 934

Db 1187 CCTCTGCTCGAGACCCCTGATTTTGAAGCTGAGGA 1222

RESULT 15

ADBI9143

ID ADBI9143 standard; cDNA; 1675 BP.

XX

AC ADBI9143;

XX

DT 20-NOV-2003 (first entry)

XX

DE Novel human secreted and transmembrane protein PRO4409 cDNA.

XX

KW Human; secreted and transmembrane protein; PRO; gene; ss;

KW Tumour necrosis factor alpha release; TNF-alpha release;

KW glucose uptake modulator; FFA uptake modulator;

KW cell proliferation stimulator; cell differentiation stimulator;

KW cell differentiation inhibitor; cytokin.

XX

OS Homo sapiens.

XX

PN US2003068796-A1.

XX

PD 10-APR-2003.

XX

PF 15-APR-2002; 2002US-00123261.

XX

PR 31-MAR-1997; 97WO-US005230.

PR 12-JUN-1998; 98WO-US012456.

PR 14-JUL-1998; 98WO-US014552.

PR 28-AUG-1998; 98WO-US017888.

PR 10-SEP-1998; 98WO-US018824.

PR 14-SEP-1998; 98WO-US019093.

PR 14-SEP-1998; 98WO-US019094.

PR 14-SEP-1998; 98WO-US019177.

PR 16-SEP-1998; 98WO-US019330.

PR 17-SEP-1998; 98WO-US019437.

PR 07-OCT-1998; 98WO-US021141.

PR 29-OCT-1998; 98WO-US022991.

PR 29-OCT-1998; 98WO-US022992.

PR 20-NOV-1998; 98WO-US024855.

PR 01-DEC-1998; 98WO-US025108.

PR 05-JAN-1999; 99WO-US000106.

PR 08-MAR-1999; 99WO-US005028.

PR 10-MAR-1999; 99WO-US005190.

PR 20-APR-1999; 99WO-US008615.

PR 14-MAY-1999; 99WO-US010733.

PR 02-JUN-1999; 99WO-US012252.

PR 01-SEP-1999; 99WO-US020111.

PR 08-SEP-1999; 99WO-US020594.

PR 13-SEP-1999; 99WO-US020944.

PR 15-SEP-1999; 99WO-US021090.

PR 15-SEP-1999; 99WO-US021547.

PR 05-OCT-1999; 99WO-US023089.

PR 29-NOV-1999; 99WO-US028214.

PR 30-NOV-1999; 99WO-US028313.

PR 30-NOV-1999; 99WO-US028409.

PR 01-DEC-1999; 99WO-US028301.

PR 01-DEC-1999; 99WO-US028634.

PR 02-DEC-1999; 99WO-US028551.

PR 02-DEC-1999; 99WO-US028564.

PR 02-DEC-1999; 99WO-US028565.

PR 16-DEC-1999; 99WO-US030095.

PR 20-DEC-1999; 99WO-US030911.

PR 20-DEC-1999; 99WO-US030999.

PR 22-DEC-1999; 99WO-US030720.

PR 30-DEC-1999; 99WO-US031243.

PR 30-DEC-1999; 99WO-US031274.

PR 05-JAN-2000; 2000WO-US000219.

PR 06-JAN-2000; 2000WO-US000277.

PR 06-JAN-2000; 2000WO-US000376.

PR 11-FEB-2000; 2000WO-US003565.

PR 18-FEB-2000; 2000WO-US004341.

PR 18-FEB-2000; 2000WO-US004342.

PR 22-FEB-2000; 2000WO-US004414.

PR 24-FEB-2000; 2000WO-US004914.

PR 24-FEB-2000; 2000WO-US005004.

PR 01-MAR-2000; 2000WO-US005601.

PR 02-MAR-2000; 2000WO-US005746.

PR 02-MAR-2000; 2000WO-US005841.

PR 10-MAR-2000; 2000WO-US006319.

PR 15-MAR-2000; 2000WO-US006884.

PR 20-MAR-2000; 2000WO-US007377.

PR 21-MAR-2000; 2000WO-US007532.

PR 30-MAR-2000; 2000WO-US008439.

PR 17-MAY-2000; 2000WO-US013705.

PR 22-MAY-2000; 2000WO-US014042.

PR 30-MAY-2000; 2000WO-US014941.

PR 02-JUN-2000; 2000WO-US015264.

PR 28-JUL-2000; 2000WO-US020710.

PR 11-AUG-2000; 2000WO-US022031.

PR 23-AUG-2000; 2000WO-US023522.

PR 24-AUG-2000; 2000WO-US023328.

PR 08-NOV-2000; 2000WO-US030952.

PR 10-NOV-2000; 2000WO-US030873.

PR 01-DEC-2000; 2000WO-US032578.

PR 20-DEC-2000; 2000US-00747259.

PR 20-DEC-2000; 2000WO-US034956.

PR 28-FEB-2001; 2001US-00796498.

PR 28-FEB-2001; 2001WO-US006520.

PR 01-MAR-2001; 2001WO-US006666.

PR 09-MAR-2001; 2001US-00802706.

PR 14-MAR-2001; 2001US-00808689.

PR 22-MAR-2001; 2001US-00816744.

PR 05-APR-2001; 2001US-00828366.

PR 10-MAY-2001; 2001US-00854208.

PR 10-MAY-2001; 2001US-00854280.

PR 18-MAY-2001; 2001US-00860216.

PR 25-MAY-2001; 2001US-00866028.

PR 25-MAY-2001; 2001US-00866034.

PR 25-MAY-2001; 2001WO-US017092.

PR 01-JUN-2001; 2001US-00872035.

PR 01-JUN-2001; 2001WO-US017800.

PR 05-JUN-2001; 2001US-00874503.

PR 14-JUN-2001; 2001US-00882836.

PR 19-JUN-2001; 2001US-00886342.

PR 20-JUN-2001; 2001WO-US019692.

PR 21-JUN-2001; 2001US-00887879.

PR 22-JUN-2001; 2001WO-US020116.

PR 29-JUN-2001; 2001WO-US021066.

PR 09-JUL-2001; 2001WO-US021735.

PR 18-JUL-2001; 2001US-00908827.

PR 06-AUG-2001; 2001US-00924419.

PR 09-AUG-2001; 2001US-00927796.

PR 16-AUG-2001; 2001US-00931836.

PR 19-DEC-2001; 2001US-00028072.

XX

FA (GETH ) GENENTECH INC.

XX

PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;

PI Geritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;

PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

XX

DR WPI; 2003-695927/66.

DR P-PSDB; ADBI9144.

XX

XX Novel secreted and transmembrane PRO polypeptides useful for stimulating the release of tumor necrosis factor alpha and detecting the presence of a tumor in a mammal.

PT

PS Claim 2; Fig 135; 660pp; English.

XX



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:40:35 ; Search time 355 Seconds  
(without alignments)  
9208.270 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacgaggaggagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/1\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/H\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/PP\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/ina/RE\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	467	25.4	1201	3	US-09-023-655-140
2	458.6	24.9	1534	3	Sequence 140, App
3	160.2	8.7	744	3	Sequence 274, App
4	160.2	8.7	1512	3	Sequence 3, Appli
5	156.6	8.5	1658	3	Sequence 1, Appli
6	75.2	4.1	7218	2	Sequence 23, Appl
7	59.4	3.2	24207	3	Sequence 14, Appl
8	56.2	3.1	7218	2	Sequence 14364, A
9	55.4	3.0	1926	3	Sequence 14, Appl
10	55.4	3.0	1931	2	Sequence 4, Appli
11	45.4	2.7	3130	3	Sequence 2, Appli
12	49.4	2.7	87734	3	Sequence 5779, Ap
13	48.6	2.6	3489	2	Sequence 17521, A
14	48.6	2.6	3489	2	Sequence 1, Appli
15	48.6	2.6	3489	3	Sequence 1, Appli
16	48.6	2.6	3489	3	Sequence 1, Appli
17	48.6	2.6	32207	2	Sequence 20, Appl
18	48.6	2.6	32207	3	Sequence 20, Appl
19	48.6	2.6	32207	3	Sequence 20, Appl
20	47.4	2.6	601	3	Sequence 91307, A
21	47.4	2.6	12695	3	Sequence 16775, A
22	46.8	2.5	289	3	Sequence 17, Appl
23	46.8	2.5	289	3	Sequence 17, Appl
24	46.8	2.5	15252	3	Sequence 13584, A

25	46.4	2.5	64309	3	US-09-949-016-14581	Sequence 14581, A
26	46.4	2.5	767677	3	US-09-949-016-12147	Sequence 12147, A
27	46.4	2.5	767677	3	US-09-949-016-17361	Sequence 17361, A
28	46	2.5	18955	3	US-09-949-016-13343	Sequence 13343, A
29	46	2.5	30678	3	US-09-949-016-12818	Sequence 12818, A
30	45	2.4	474	3	US-09-621-976-18033	Sequence 18033, A
31	44.8	2.4	3053	3	US-10-104-047-958	Sequence 958, App
32	44.6	2.4	3319	2	US-08-006-676B-2	Sequence 2, Appli
33	44.6	2.4	3319	2	US-08-282-845-1	Sequence 1, Appli
34	44.6	2.4	3319	2	US-08-428-414A-4	Sequence 4, Appli
35	44.6	2.4	3319	6	PCT-US94-00324-2	Sequence 2, Appli
36	44.4	2.4	1926	3	US-09-249-585A-2	Sequence 2, Appli
37	44.4	2.4	1926	3	US-09-410-399-3	Sequence 3, Appli
38	44.4	2.4	2580	3	US-09-050-863-2	Sequence 2, Appli
39	44.4	2.4	2580	3	US-09-359-081-2	Sequence 2, Appli
40	44.4	2.4	5452	2	US-09-130-114-1	Sequence 1, Appli
41	44.4	2.4	8705	3	US-09-647-344A-14	Sequence 14, Appli
42	44.4	2.4	9600	3	US-08-910-647-1	Sequence 1, Appli
43	44.4	2.4	9600	3	US-09-620-925-1	Sequence 1, Appli
44	44.4	2.4	10596	2	US-07-884-811-15	Sequence 15, Appli
45	44.4	2.4	10596	2	US-07-885-971-15	Sequence 15, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-023-655-140  
; Sequence 140, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 140:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: THPINOB01  
; CLONE: 030137

US-09-023-655-140

```
Query Match      25.4%; Score 467; DB 3; Length 1201;
Best Local Similarity 79.0%; Pred. No. 4.1e-109;
Matches 612; Conservative 0; Mismatches 95; Indels 68; Gaps 2;

QY 228 GCTGAAGTCGGCTTTTGAGGAAACGGGAAAGACCAAGGAAGTGATTTGACACCGGTATGG 287
Db 1 GCTGAAGTCAGCCTTTTGAGGAAACCGGCAAGACCAAGGAGGTGATTTGGCACGGGCTATGG 60

QY 288 CATCTGGACGGGAAAGGGCTCTGGAGTCAAGTACACCAAGTC----- 329
Db 61 CATCTGGACCAAGAGGGCTCTGGAGTCAAAATACAAAGTCCATTTTCAGATCCCCCAGA 120

QY 330 -----GGACTTACGGTTAATT 345
Db 121 CCAGATGACCTATCTTCTCCAGCTCTGAGTCACTTCCCATGCGGACTTCGGGTAAATC 180

QY 346 GAAGTCACTGAGACCATTTTGAAGAGGCTTCTGGACTACAGCTGCACAAAGGAGGAGT 405
Db 181 GAAGTCACTGAGACCATTTTGAAGAGGCTTCTGGATATAGCTGCACAAAGGAGGAGT 240

QY 406 GCGAGCAACCGGTTTCCCAAGGGTATGTCGAGACCTTTTGAGACGCTGCACAACTAGTC 465
Db 241 GCGAGCAATCGATTTGCCAAGGGCATGTCAGAGACCTTTGAGACATTACACAACTGGTA 300

QY 466 CACAAGGGGTCAAGGTGGTATGATATCCCTATAGCTGTGGAACGAGACCTCAGCA 525
Db 301 CACAAGGGGTCAAGGTGGTATGATGACATCCCTATAGCTGTGGAACGAGACTTCGTGA 360

QY 526 GAGGTGGCTGACCTCAAGAAGCAGTGTGACCTGCTGGTGGAGAGTGTGAAGAGGTGATT 585
Db 361 GAGGTGGCTGACCTCAAGAAGCAGTGTGATGTGCTGGTGGAGAGTGTGAAGAGGTGATC 420

QY 586 GAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTCTGTGCCAACCA 645
Db 421 GAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTCTGTGCCAACCA 480

QY 646 GTGCTGAAGGGAAGGACACAGTTGCTCTAGCAGAGCGGTGCTGGCAAGAGGGGAC 705
Db 481 GTGCTGAAGGGAAGGACACAGTTGCTCTGCGAGAGCAGTGGTCCGGCAAGAGGGAGAC 540

QY 706 ATAGCTCTCCCTGGGAGGGAAGAAATCCAAAGAAAGCGCAGCGGAGTCAAG-----GGC 759
Db 541 ACAGTGGCTCTGGGAGGGAAGAGTCCAAAGAAAGACAGACAGAGGCCCAAGGACGAGGC 600

QY 760 TCCTTCAGTGGCAGCAGCAAGCAGAGGAAGAACTGGGGGGCTTGGGGGAGGATGCCAAC 819
Db 601 GCGAGGAGTAGCAGCAGCAAAACAAAGGAAGAGCTGGGTGGCTTTGAGGGGAGACCCACG 660

QY 820 GCGAGGAGGAGGAGGTGTGAGNAGCANTGCCCCCTCCACACAGAGCCCCCTGTATGAG 879
Db 661 CCCGAGGAGGATGAGGGGATCCAGAAGGCATCCCTCTCTCACACACAGCCCCCTGTATG 720

QY 880 CTGTGAGCCCAAGCTTAGTGTCTTCAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 721 CTCTGAGCCCAAGCAGCATCTCTGTCTGAGACCCCTGATTTGAAGCTGAGGA 775
```

## RESULT 2

```
US-09-023-655-274
; Sequence 274, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
```

```
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA: US/09/023,655
; APPLICATION NUMBER: 09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 274:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1534 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTNOT07
; CLONE: 1298861
; US-09-023-655-274
```

```
Query Match      24.9%; Score 458.6; DB 3; Length 1534;
Best Local Similarity 84.8%; Pred. No. 6.3e-107;
Matches 525; Conservative 0; Mismatches 92; Indels 2; Gaps 2;

QY 5 CGAGGAGGAAGCGCCCGGGTCCGCTCTGCTCGGTCCGGTCCGGCCATGGAGTCCA 64
Db 660 CGCGGGAGGAGGAACCCCGGGTCTTTAGGGTCCGGGCCCGGGCCATGATTCAA 719

QY 65 TGTCTGAGCTCGCGCCCGGTCCCTTATTTCTTTGCTGCTGCTGCTTC-CGCTGCTG 123
Db 720 TGCTTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCTGCTGCTGCTGCTGCTGCTG 779

QY 124 CTCCTTCTCTCCCGAAGTAGGCCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTG 183
Db 780 NNNNNNNNGCCCGGAGCTGGGCCCGAGCCGAGCTGAGGAGAACGACTGGGTT 839

QY 184 CGATTGCCAGCAATCGGAAGTGTGCAAGTATGTTGCTGTGGAGCTG-AGATCGGCTTT 242
Db 840 CGCTGCGCCAGCAATCGAAGTGTGTAATATGTTGCTGTGGAGCTGTAAGTCAGCCTT 899

QY 243 TGAGAAACGGGAAAGACCAAGGAAGTATGACACCGGCTATGGCATCTCGGACGGAA 302
Db 900 TGAGAAACCGGCAAGACCAAGGAGGTATTTGGCACCGGCTATGGCATCTCGGACGAA 959

QY 303 GGGCTCTGGAGTCAAGTACACCAAGTCGAGTTCGAGTAAATGGAAGTCACTGAGACCAT 362
Db 960 GGCCTCTGGAGTCAATATACCAAGTCGAGTTCGAGTAAATGGAAGTCACTGAGACCAT 1019

QY 363 TTGCAAGAGGCTTCTGGACTACAGCTGCACAAAGGAGGACTGGCAGCAACCGGTTTGC 422
Db 1020 TTGCAAGAGGCTTCTGGATTTATAGCTTGCACAAAGGAGGACCGGAGCAATTCGATTCG 1079

QY 423 CAAGGATGTGCGAGAGCTTTTGAGAGCTCTGCACAACTAGTCCACAAAGGGGTCAAGGT 482
Db 1080 CAAGGATGTGAGAGCTTTTGAGATTTACACAACTGTGTACACAAAGGGGTCAAGGT 1139

QY 483 GGTGATGATATCCCTTATGAGCTGTGGAACGAGACTTCAGCAGAGGTGGCTGACCTCAA 542
Db 483 GGTGATGATATCCCTTATGAGCTGTGGAACGAGACTTCAGCAGAGGTGGCTGACCTCAA 542
```

Db 1140 GGTGATGACATCCCTATGAGCTGTGGAACGAGACTTCTGACAGGTGGCTGACCTCAA 1199  
Qy 543 GAAGCAGTGTGACGTGCTGGTGGAGAGTTTGAACAGGTGATTGAGGACTGGTACAGGAA 602  
Db 1200 GAAGCAGTGTGATGTGCTGGTGGAGAGTTTGAAGGAGGTGATCGAGGACTGGTACAGGAA 1259  
Qy 603 CCACCAGGAGGAGACCTG 621  
Db 1260 CCACCAGGAGGAGACCTG 1278

## RESULT 3

US-09-163-285-3  
; Sequence 3, Application US/09163285  
; Patent No. 6204013  
; GENERAL INFORMATION:  
; APPLICANT: Khodadoust, Mehran  
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/163,285  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/090,398  
; FILING DATE: June 24, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: MMI-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 744 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..744  
US-09-163-285-3

Query Match 8.7%; Score 160.2; DB 3; Length 744;  
Best Local Similarity 57.3%; Pred. No. 6.9e-31;  
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;  
Qy 162 GCCTCAGGAGACCGGACTGGTGGTGGGATTCGCCAGCAATGCGAAGTGTGCAAGTATGTTGC 221  
Db 75 GGAGGAGACGNTGACACAGAACGCTTCCCGAGCAATGCGAAGTGTGTAAGTGTCTGAG 134  
Qy 222 TGTGAGCTGAAGTCCGCTTTTGGAGAAACGGGAAAGACCAAGGAAGTGTGACACCGG 281  
Db 135 CACAGACTACAGGGCGGAACAGTGCACCGGTGCATCTCAGAGAGTGTGGAGCTGGG 194  
Qy 282 CTATGGCATCTTGGAGCGGAGGGCTCTGGAGTCAAG---TACACCAAGTGGACTTACG 338  
Db 195 GCAGGTGCTGGATACAGGCAAGAGGAGAGACACGTGCTTTACAGCGTTTCAGAGACAAG 254

Qy 339 GTTAATTGAGTCACTGAGACCATTTGCAAGAGGCTTCTGGACTACAGCCTGCACAGGA 398  
Db 255 GCTGGAAGAGCGCTTAGAGAAATTTATGTGAGCGGATCCTGGACTATAGTGTTCACGCTGA 314  
Qy 399 GAGGACTGGCAGCAACCGGTTTGCACAGGGTATGTCCGAGACCTTTGAGACGCTGCACAA 458  
Db 315 GCGCAAGGGCTCACTGAGATATGCCAAGGGTCAGAGTCAGACCATGGCAACACTGAAAGG 374  
Qy 459 CCTAGTCCAAAAGGGGTCAAGGGTGTGATGATATCCCTATGAGCTGTGGAACAGAGAC 518  
Db 375 CCTAGTGCAGAAAGGGGTGAAGGTGGATCTCGGGATCCCTCTGGAGCTTTGGGATGAGCC 434  
Qy 519 CTCAGCAGAGTGGCTGACCTCAAGAACCATGTGACGTGCTGGTGGAGAGTTTGAAGA 578  
Db 435 CAGCGTGGAGGTCACTACCTCAAGAACCATGTGACGTGCTGGAGAGTTTGAAGA 494  
Qy 579 GGTGATTGAGACTGGTACAGAAACCCACGAGGAGAACCTGACTGAATTCCTCTGTGC 638  
Db 495 CATTGTGGAGACTGGTACTTCCACCATCAGAGAGCGCCCTACAAAATTTCTCTGTGA 554  
Qy 639 CAACCACTGCTGAAGGGAAGGACACGAGTTGCTAGCAGAGCGGTGCTGGCAAGAA 698  
Db 555 AGGTCACTGCTCCAGCTGCTGAAACTGCATGCTACAGGAAACTTGGACTGGAAAGGA 614  
Qy 699 G 699  
Db 615 G 615

## RESULT 4

US-09-163-285-1  
; Sequence 1, Application US/09163285  
; Patent No. 6204013  
; GENERAL INFORMATION:  
; APPLICANT: Khodadoust, Mehran  
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/163,285  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/090,398  
; FILING DATE: June 24, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: MMI-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1512 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:

```
; NAME/KEY: CDS
; LOCATION: 125...868
US-09-163-285-1

Query Match      8.7%; Score 160.2; DB 3; Length 1512;
Best Local Similarity 57.3%; Pred. No. 9e-31;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

Qy 162 GGCTGAGGAGACCGACTGGTGGATTCGCCAGCAAAATGCCAAGTGTGCAAGTATGTTGC 221
Db 199 GGAGGAGGAGCATGATCACAGAACCGCTTGCCAGCAAAATGCCAAGTGTGTAAGTGC 258
Qy 222 TGTGAGCTGAAGTCGGCTTTTGAGGAAACGGGAAAGACCAAGGAAGTGTGACACGG 281
Db 259 CACAGAGCTACAGGGGAACTGAGTCGCACCGTGCATCTCGAAGGTGCTGGAGCTGG 318
Qy 282 CTATGGCATCTGGACGGGAAGGGCTCTGGAGTCAAG---TACCAAGTCCGACTTAAG 338
Db 319 GCAGGTGCTGGATACAGGCAAGAGGAGAGACACGTGCTTTACAGCGTTTCAGAGACAAG 378
Qy 339 GTTAATTGAAGTCACTGAGACCATTTGCAAGAGGCTTCTGCACTACAGCCTGCAACAGGA 398
Db 379 GCTGGAAGAGCCCTTAGAGAAATTTATGTGACGGGATCCTGGACTATAGTGTTCACGCTGA 438
Qy 399 GAGGACTGGCAGCAACCGGTTTGGCAAGGGTATGTCCGAGACCTTTGAGAGCGCTGCACAA 458
Db 439 CGCAAGGGCTCACTGAGATATGCCAAGGTTACAGTCAAGCCATGGCAACACTGAAAGG 498
Qy 459 CCTAGTCCACAAAGGGGTCAAGGTGGTGATGGATATCCCTATGAGCTGTGGAAACGAGAC 518
Db 499 CCTAGTCAGAAAGGGGTGAAGTGGATCTCGGGATCCCTCTGGAGCTTTGGGATGAGCC 558
Qy 519 CTCAGCAGAGTGGTGAACCTCAAGAACAGTGTGACGTGCTGGTGGAAAGTTTGAAGA 578
Db 559 CAGCGTGGAGGTGCATACATACCTCAAGAACAGTGTGAGACCATGTTTGGAGGAGTTTGAAGA 618
Qy 579 GGTGATTGAGACTGGTACAGAACCCACGAGGAAGACCTGACTGAATTCCTCTGTGC 638
Db 619 CATTGGGGAGCTGGTATCTCCACCATCAGAGCAGCCCTTACAAATTTTCTGTGA 678
Qy 639 CAACCATGCTGTAAGGGAAGGACACGAGTTGCTTAGCAGGCGGTGCTGCGCAAGAA 698
Db 679 AGGTATGCTGCTCCAGCTGCTGAAACTGCATGCTCAGAGGAACCTTGGACTGGAAAGGA 738
Qy 699 G 699
Db 739 G 739

RESULT 5
US-09-482-273-23/c
; Sequence 23, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030PL
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1658
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-482-273-23
Query Match      8.5%; Score 156.6; DB 3; Length 1658;
Best Local Similarity 76.2%; Pred. No. 7.8e-30;
Matches 208; Conservative 0; Mismatches 59; Indels 6; Gaps 1;

Qy 668 GTTCCCTTAGCAGACGGTGTCTGGCAAGAAGGGGGACATAGCCTCCCTGGGAGGGAAGA 727
Db 308 GTTCCCTGGCAGACAGTGTCCGGCAAGAAGGGAGACACACAGCTGCCCTGGGAGGAAGA 249
Qy 728 AATCAAGAAGAACCGCAGCGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCAAGC 781
Db 248 AGTCCAAGAAGAACGAGCAGCGGCGCAAGGCGCAGCGGAGGAGTAGCAGCAGCAAAC 189
Qy 782 AGAGGAAGGAAGTGGGGGGCTGGGGAGGATGCAACGCCGAGGAGGAGGAGGTGTGC 841
Db 188 AAAGGAAGGAGCTGGGTGGCTTGAGGAGACCCAGCCCGAGGAGATGAGGCGATCC 129
Qy 842 AGAAGGATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCCCAGCTTAGTGTCC 901
Db 128 AGAAGGATCCCCCTCTCACACACAGCCCCCTGATGAGCTGTGAGCCCCAGCATCCT 69
Qy 902 TTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 68 CTGTCTGAGACCCCTGATTTTGAAGCTGAGGA 36

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
```



```
Query Match      3.2%; Score 59.4; DB 3; Length 24207;
Best Local Similarity 67.2%; Pred. No. 0.00014;
Matches 84; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```



; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 5779  
 ; LENGTH: 3130  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-949-016-17521

Query Match 2.7%; Score 49.4; DB 3; Length 3130;  
 Best Local Similarity 47.3%; Pred. No. 0.023;  
 Matches 181; Conservative 0; Mismatches 201; Indels 1; Gaps 1;  
 QY 189 GCCAGCAAAATGCGAAGTGTGCA-AGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTGAGG 247  
 DB 1745 GCGCGGCCCAACGAGAAGTACGACAGGAGGTGGCGGGCTTGAAGGACAAGGTTTCAGCAGG 1804  
 QY 248 AAACGGGAAAGACCAAGAGTGTGACACCGGCTATGGCATCTGGACCGGAAGGGCT 307  
 DB 1805 CCACAGGAGAACATGGGGCTAATGGAACTGGAATCCAGCTGACCTCGCTGGCT 1864  
 QY 308 CTGGAGTCAAGTACCAAGTCGGACTTACGGTTAATTGAAGTCACTGAGACCACTTTGCA 367  
 DB 1865 CGGACCACCAAGAGTCCCTGGAGGACCTCAAGCCACCTGAACCTCGGGCCAGCGGCC 1924  
 QY 368 AGAGGCTTCTGCACTACAGCTGCAACAGGAGGAGCTGGCAGCAACCGGTTTTCGAAGG 427  
 DB 1925 AGCAGAGGAGATCGCGGAGCTGAAGGAGTGTGAGGAGGATCAAGATGGAGCACCAGC 1984  
 QY 428 GTATGTCGAGACCTTTCAGAGCTGCAACACCTAGTCCAAAGGGGTCAAGTGTGA 487  
 DB 1985 TGGAGCTGGGTAACCTTTCAGGCGCAAGCATGACCTGGAGACCGCATGCACGTGAAGGAGA 2044  
 QY 488 TGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAGC 547  
 DB 2045 AGGAGGCGCTCGAGAGAAAGTGCAGGAGGCCCGCAGGAGGAGTGGCTGGCTGCAGCGGC 2104  
 QY 548 AGTGTGAGCTGCTGTGGAAGAG 570  
 DB 2105 ACTGCGGGGCCAGCTGAGGTG 2127

RESULT 12  
 US-09-949-016-17521  
 ; Sequence 17521, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307  
 ; CURRENT FILING DATE: 2000-04-14  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; PRIOR FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 17521  
 ; LENGTH: 87734  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-949-016-17521

Query Match 2.7%; Score 49.4; DB 3; Length 87734;

Best Local Similarity 47.3%; Pred. No. 0.082;  
 Matches 181; Conservative 0; Mismatches 201; Indels 1; Gaps 1;  
 QY 189 GCCAGCAAAATGCGAAGTGTGCA-AGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTGAGG 247  
 DB 60303 GCGCGGCCCAACGAGAAGTACGACAGGAGGTGGCGGGCTTGAAGGACAAGGTTTCAGCAGG 60362  
 QY 248 AAACGGGAAAGACCAAGAGTGTGACACCGGCTATGGCATCTGGACCGGGAAGGGCT 307  
 DB 60363 CCACAGGAGAACATGGGGCTAATGGAACTGGAATCCAAAGCTGCACTCGCTGGCT 60422  
 QY 308 CTGGAGTCAAGTACCAAGTCGGACTTACGGTTAATTGAAGTCACTGAGACCACTTTGCA 367  
 DB 60423 CGGACCACCAAGAGTCCCTGGAGGACCTCAAGCCACCTGAACTCGGGCCAGCGGCC 60482  
 QY 368 AGAGGCTTCTGCACTACAGCTGCAACAGGAGGAGCTGGCAGCAACCGGTTTTCGAAGG 427  
 DB 60483 AGCAGAGGAGATCGCGGAGCTGAAGGAGTGTGAGGAGGATCAAGATGAGCACCAGC 60542  
 QY 428 GTATGTCGAGACCTTTCAGAGCTGCAACACCTAGTCCCAAAAGGGGTCAAGTGTGA 487  
 DB 60543 TGGAGCTGGGTAACCTTTCAGGCGCAAGCATGACCTGGAGACCGCCATGCACGTGAAGGAGA 60602  
 QY 488 TGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAGC 547  
 DB 60603 AGGAGGCGCTCGAGAGAGAGCTGAGGAGGCCAGGAGGAGTGGCTGGCTGCAGCGGC 60662  
 QY 548 AGTGTGAGCTGCTGTGGAAGAG 570  
 DB 60663 ACTGCGGGGCCAGCTGAGGTG 60685

RESULT 13  
 US-08-728-323A-1  
 ; Sequence 1, Application US/08728323A  
 ; Patent No. 5948676  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chang, Yuan  
 ; APPLICANT: Bohenzky, Roy A.  
 ; APPLICANT: Russo, James J.  
 ; APPLICANT: Edelman, Isidore S.  
 ; APPLICANT: Moore, Patrick S.  
 ; TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
 ; Sarcoma-Associated Herpesvirus, DNA  
 ; TITLE OF INVENTION: Encoding Same And Uses Thereof  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/728,323A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-278-0400  
 ; TELEFAX: 212-391-0525  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3489 base pairs  
 ; TYPE: nucleic acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1

Query Match          2.6%; Score 48.6; DB 2; Length 3489;
Best Local Similarity 48.4%; Pred. No. 0.038;
Matches 194; Conservative 0; Mismatches 204; Indels 3; Gaps 2;

Qy 446 AGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGGTATGATATCCCTTATGAGC 505
Db 2243 AGCAGCAGGATGAACAGGAGCAGCAGGAGGAGCAGCAGGAGGAGCAGGAGCAGG 2302
Qy 506 TGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCAGTGTGACGTGCTGTGG 565
Db 2303 AGTTAGAGGAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGC 2362
Qy 566 AAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGACCAACCAGGAGGAACCTGACTG 625
Db 2363 AGGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAGCAGGAGTTAGAGGAGCAGGAGC 2422
Qy 566 AAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGACCAACCAGGAGGAACCTGACTG 625
Db 2363 AGGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAGCAGGAGTTAGAGGAGCAGGAGC 2422
Qy 626 AATTCCTCTGTGCCAACCAACGCTGTGAAGGAAAGGACACGAGTTGCTTAGCAGAGCGGT 685
Db 2423 AGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGC--AGGAGCAGGAGTTAGAGGA 2480
Qy 686 GGTCTGGCAAGAGGGGGACATAGCCTCCCTGGGAGGGAAGAAATCCAAAGAAAGCGCA 745
Db 2481 GCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTGGAAGAGCA 2540
Qy 746 CGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAGCAGCAGGAGGAAGAACTGGGGGCGCTGG 805
Db 2541 AGAGCAGGAGTGGAAAGAGCAAGAGCAGGAGCAGGAGCAAGAGC--AGGAATTAGAGGAGGTG 2599
Qy 806 GGGAGGATGCCAACCGCCGAGGAGGAGGAGGTGTGCAGAAAG 846
Db 2600 AGGAGCAAGACGAGGAGCAGGAGGAGGAGGAGGAGGAG 2640

```

```

RESULT 14
US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 632792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Balleskas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match          2.6%; Score 48.6; DB 3; Length 3489;
Best Local Similarity 48.4%; Pred. No. 0.038;
Matches 194; Conservative 0; Mismatches 204; Indels 3; Gaps 2;

Qy 446 AGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGGTATGATATCCCTTATGAGC 505
Db 2243 AGCAGCAGGATGAACAGGAGCAGCAGGAGGAGCAGCAGGAGGAGCAGGAGCAGG 2302
Qy 506 TGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCAGTGTGACGTGCTGTGG 565

```

```

Db 2303 AGTTAGAGGAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGC 2362
Qy 566 AAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCAACCAGGAGGAACCTGACTG 625
Db 2363 AGGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAGCAGGAGTTAGAGGAGCAGGAG 2422
Qy 626 AATTCCTCTGTGCCAACCAACGCTGTGAAGGAAAGGACACGAGTTGCTTAGCAGAGCGGT 685
Db 2423 AGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGC--AGGAGCAGGAGTTAGAGGA 2480
Qy 686 GGTCTGGCAAGAGGGGGACATAGCCTCCCTGGGAGGGAAGAAATCCAAAGAAAGCGCA 745
Db 2481 GCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTGGAAGAGCA 2540
Qy 746 CGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAGCAGCAGGAGGAAGAACTGGGGGCGCTGG 805
Db 2541 AGAGCAGGAGTGGAAAGAGCAAGAGCAGGAGCAGGAGCAAGAGC--AGGAATTAGAGGAGGTG 2599
Qy 806 GGGAGGATGCCAACCGCCGAGGAGGAGGAGGTGTGCAGAAAG 846
Db 2600 AGGAGCAAGACGAGGAGCAGGAGGAGGAGGAGGAGGAG 2640

RESULT 15
US-09-410-399-1
; Sequence 1, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

Query Match          2.6%; Score 48.6; DB 3; Length 3489;
Best Local Similarity 48.4%; Pred. No. 0.038;
Matches 194; Conservative 0; Mismatches 204; Indels 3; Gaps 2;

Qy 446 AGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGGTATGATATCCCTTATGAGC 505
Db 2243 AGCAGCAGGATGAACAGGAGCAGCAGGAGGAGCAGGAGCAGGAGGAGCAGGAGCAGG 2302
Qy 506 TGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCAGTGTGACGTGCTGTGG 565
Db 2303 AGTTAGAGGAGCAGGAGCAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGC 2362
Qy 566 AAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCAACCAGGAGGAACCTGACTG 625
Db 2363 AGGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAGCAGGAGTTAGAGGAGCAGGAG 2422
Qy 626 AATTCCTCTGTGCCAACCAACGCTGTGAAGGAAAGGACACGAGTTGCTTAGCAGAGCGGT 685
Db 2423 AGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGC--AGGAGCAGGAGTTAGAGGA 2480
Qy 686 GGTCTGGCAAGAGGGGGACATAGCCTCCCTGGGAGGGAAGAAATCCAAAGAAAGCGCA 745
Db 2481 GCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTGGAAGAGCA 2540
Qy 746 CGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAGCAGCAGGAGGAAGAACTGGGGGCGCTGG 805
Db 2541 AGAGCAGGAGTGGAAAGAGCAAGAGCAGGAGCAGGAGCAAGAGC--AGGAATTAGAGGAGGTG 2599
Qy 806 GGGAGGATGCCAACCGCCGAGGAGGAGGAGGTGTGCAGAAAG 846

```

**Fri Dec 30 11:17:44 2005**

Db 2600 AGGAGCAAGAGCAGGAGCAGGAGCAGGAGCAGGAGGAGCAGGAG 2640

Search completed: December 30, 2005, 03:44:24  
Job time : 359 secs

**THIS PAGE BLANK (USPTO)**



261	AGGAAA	CCGGCAAGAC	CAAGAGGTGAT	TGGCACA	CGGGCTAT	TGGCAT	TCTGGAC	CAAGAAGG	320
305	GCTCTG	GAGTCAAG	TACACCA	AGTCGGA	CTTACGGT	TAAATG	GAAGTCA	CTGAGACCA	364
321	CCTCTG	GAGTCAAA	TATACAC	CAAGTCG	GAATTC	CGGTTAAT	CGAAGTCA	CTGAGACCA	380
365	GCAAGAG	GCTTCTG	GACTTAC	AGCCTG	GCACAG	GAGAGG	AGGACTTGG	CAGCAACCG	424
381	GCAAGAG	GCTCCTG	GAATAT	TAGCCT	TGCAACAG	GAGAGG	ACCGG	CAGCAATTC	440
425	AGGCTAT	GTCGGAG	ACCTTTG	AGACG	CGCTGC	CAACCT	TAGTCC	CAAAAGGGG	484
441	AGGCA	TGTCAG	AGACCTT	TGAG	CAATTAC	CAACCT	TGTTAC	CAAAAGGGG	500
485	TGATGGA	TATCCCT	TATG	AGCTGT	TGGAA	CGAG	ACCTCAG	CAGAGTGG	544
501	TGATGGA	CATCCCT	TATG	AGCTGT	TGGAA	CGAG	ACTTCTG	CAGAGTGG	560
545	AGCAGT	GTGAC	GTCTG	TGTG	GAAGAT	TTGA	AGAGTGT	ATTCAG	604
561	AGCAGT	GTGAT	GTCTG	TGTG	GAAGAT	TTGA	AGAGTGT	ATTCAG	620
605	ACCAGG	AGGAAG	ACTG	ATCT	CTGTG	CCAA	CCACG	TGTAAGG	664
621	ACCAGG	AGGAAG	ACTG	ATCT	CTGTG	CCAA	CCACG	TGTAAGG	680
665	CGAGT	GCCTAG	CAGAG	CGGTGG	TCTG	CCAAAG	AGGGG	ACATAG	724
681	CCAGT	GTGCTG	GCAG	AGCAG	TGGT	CCGAC	AGAGG	GAGACAC	740
725	AGAAAT	CCAGA	AGAC	CGCAG	CGAGT	CAAG	-----GG	TCTCCAGT	778
741	AGAAAT	CCAGA	AGAC	CGCAG	CGAGT	CAAG	-----GG	TCTCCAGT	800
779	AGCAGAG	AAGGA	ACTGG	GGGG	CCCTGG	GGGAGG	ATGCC	AA	838
801	AACAAAG	GAAGG	AGCTGG	TGGT	GGCCCT	TGAG	GGGAG	ACCC	860
839	TGCAGAG	GGCAT	CGCC	CCCT	TCC	CAC	CAGC	CC	898
861	TCCAGAG	GGCAT	CGCC	CCCT	TCC	CAC	CAGC	CC	920
899	TCCTTGA	ATCAAG	ACCC	CTG	ACTT	CAG	AGCT	TTGG	934
921	CCTCTGT	CCTCAG	ACCC	CTG	ACTT	CAG	AGCT	TTGG	956

RESULT 2  
US-10-431-096-139  
Sequence 139, Application US/10431096  
Publication No. US2004008696A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB  
TITLE OF INVENTION: PATHWAY  
FILE REFERENCE: D0108A CIP  
CURRENT APPLICATION NUMBER: US/10/431,096  
CURRENT FILING DATE: 2003-05-07  
PRIOR APPLICATION NUMBER: US 60/284,962  
PRIOR FILING DATE: 2001-04-19  
PRIOR APPLICATION NUMBER: US 10/126,103  
PRIOR FILING DATE: 2002-04-19  
PRIOR APPLICATION NUMBER: US 60/286,645  
PRIOR FILING DATE: 2001-04-26  
PRIOR APPLICATION NUMBER: US 60/346,986  
PRIOR FILING DATE: 2002-01-09  
NUMBER OF SEQ ID NOS: 307  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 139  
LENGTH: 1378  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-431-096-139

Query Match	37.6%	Score	691.6	DB	7	Length	1378
Best Local Similarity	84.5%	Pred. No.	6.6e-197				
Matches	791	Conservative	0	Mismatches	137	Indels	6
Gaps	1						
QY	5	CGAGGGAGGAAGCGCGCGCGGCTCGCTCTGCTCTGGGTCGGGCTGGAGTCCA	64				
DB	21	CGCGGAGGAGGAACCCGCCGCTCTTTAGGGTCCGGGCCCGGCGCGGCGCATGATTCAA	80				
QY	65	TGCTGAGCTCGCGCCCCCGCTGCTCTTATTTCTTTGCTGCTGCTCTGCTCTGCTGCTG	124				
DB	81	TGCTGAGCGCGCTCCCGCTGCTCTTCTGCTCTTCTCCCTTTGCTGCTGCTGCTGCTG	140				
QY	125	TCCTTCTGCGCGAAGCTAGGCCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGC	184				
DB	141	TGCTGCCGGGCCCGGAGCTGGGCCGAGCCGAGCCCGGAGCTGAGGAGAACGACTGGGTTC	200				
QY	185	GATTGCCCCAGCAATGCAAGTGTGCAAGTATGTTGCTGTGGAGCTGAACTCGGCTTTTG	244				
DB	201	GCCTGCCCGCAATGCGAAGTGTGTAAATATGTTGCTGTGGAGCTGMACTCAGCCCTTG	260				
QY	245	AGGAAACGGGAAGAACCAAGGAAGTATGCACACCGGCTATATGCATCTCTGGACGGGAAG	304				
DB	261	AGGAAACCGGCAAGCAACCAAGAGGTGATTGGCAACGGGCTATGGCATCTCTGGACGAAAG	320				
QY	305	GCTCTGAGTCAAGTACACCAAGTCGGACTTACGGTTTATTTGAAGTCACTGAGACCATTT	364				
DB	321	CCTCTGGAGTCAAAATACCAAGTCGGACTTGGCGGTTTATCGAAGTCACTGAGACCATTT	380				
QY	365	GCAAGAGGCTTCGACATACAGCTGCACAAAGAGAGAGACTGCGCAGCAACCGGTTTGCCA	424				
DB	381	GCAAGAGGCTCCTGGATTATAGCCTGCACAAAGAGAGGACCGGCAGCAATCGATTGGCCA	440				
QY	425	AGGTTATGTCGAGACCTTTTGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGG	484				
DB	441	AGGTCATGTCAGAGACCTTTTGACACATTAACAACCTGGTGACAAAGGGGTCAAGGTGG	500				
QY	485	TGATGGATATCCCTATAGCTGTGGAACAGAGACCTCAGCAGAGGTGGCTGCCTCAAGA	544				
DB	501	TGATGGACATCCCCCTATGAGCTGTGGAACAGAGACTTCTGCAGAGGTGGCTGACTCAAGA	560				
QY	545	AGCAGTGTGAGTCTGGTGGAAAGTGTGAAGAGGTGATTGAGGACTGGTACAGGAACC	604				
DB	561	AGCAGTGTGATGTGCTGTGGAAGTGTTCAGGAGGTGATCGAGGACTGGTACAGGAACC	620				
QY	605	ACCAAGGAGAGACCTGACTGAATTCCTCTGTGCCAACACAGTCTGGAAGGGAAGGACA	664				
DB	621	ACCAAGGAGAGACCTGACTGAATTCCTCTGTGCCAACACAGTCTGGAAGGGAAGGACA	680				
QY	665	CGAGTCTCCTAGCAGACGGGTGTCTGGCAAGAGGGGACATAGCTCCCTCGGAGGGA	724				
DB	681	CCAGTGTCTGGCAGACAGTGTCTCGCAAGAGGGAGACACAGTGTCTCTGGAGGGA	740				
QY	725	AGAAATCCAAAGAAAGCGCAGCGAGTCAAG-----GGCTCCTCCAGTGGCAGCAGCA	778				
DB	741	AGAAATCCAAAGAAAGCAGCAGCGGACCAAGGCAGCAGCGCGCAGGATAGCAGCAGCA	800				
QY	779	AGCAGAGGAGGAACCTGGGGGGCTGGGGAGGATGCCAACCGCGAGGAGGAGGGGTG	838				
DB	801	AACAAAGGGAAGGAGCTGGGTGGCTTGAGGGAAGCCCCAGCCCCGAGGAGGATGAGGGCA	860				
QY	839	TGCAGAAGGATCGCCCTCCCAACACAGCCCCCTGATGAGTGTGAGCCCGCATAGTG	898				
DB	861	TCCAGAAGGATCCCTCTCACACAGCCCCCTGATGAGCTCTGAGCCACCACCGCAT	920				
QY	899	TCCTTTGAATCAAGACCCCTGACTTCAGAGCTTGGA	934				
DB	921	CCTCTGTCTGAGACCCCTGATTTGAGCTGAGGA	956				

RESULT 3  
US-10-028-072-135  
; Sequence 135, Application US/10028072



; Publication No. US20030004311A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Laureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang  
 ; TITLE OF INVENTION:  
 ; FILE REFERENCE:  
 ; CURRENT APPLICATION NUMBER: US/10/028,072  
 ; CURRENT FILING DATE: 2001-12-19  
 ; PRIOR APPLICATION NUMBER: 60/049911  
 ; PRIOR FILING DATE: 1997-06-18  
 ; PRIOR APPLICATION NUMBER: 60/056974  
 ; PRIOR FILING DATE: 1997-08-26  
 ; PRIOR APPLICATION NUMBER: 60/059113  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059115  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059117  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059122  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059184  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059263  
 ; PRIOR FILING DATE: 1997-09-18  
 ; PRIOR APPLICATION NUMBER: 60/059352  
 ; PRIOR FILING DATE: 1997-09-19  
 ; PRIOR APPLICATION NUMBER: 60/059588  
 ; PRIOR FILING DATE: 1997-09-19  
 ; PRIOR APPLICATION NUMBER: 60/059836  
 ; PRIOR FILING DATE: 1997-09-24  
 ; PRIOR APPLICATION NUMBER: 60/062250  
 ; PRIOR FILING DATE: 1997-10-17  
 ; PRIOR APPLICATION NUMBER: 60/062285  
 ; PRIOR FILING DATE: 1997-10-17  
 ; PRIOR APPLICATION NUMBER: 60/062287  
 ; PRIOR FILING DATE: 1997-10-17  
 ; PRIOR APPLICATION NUMBER: 60/062814  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/062816  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/063045  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/063082  
 ; PRIOR FILING DATE: 1997-10-31  
 ; PRIOR APPLICATION NUMBER: 60/063127  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/063327  
 ; PRIOR FILING DATE: 1997-10-27  
 ; PRIOR APPLICATION NUMBER: 60/063329  
 ; PRIOR FILING DATE: 1997-10-27  
 ; PRIOR APPLICATION NUMBER: 60/063550  
 ; PRIOR FILING DATE: 1997-10-28  
 ; PRIOR APPLICATION NUMBER: 60/063561  
 ; PRIOR FILING DATE: 1997-10-28  
 ; PRIOR APPLICATION NUMBER: 60/063704  
 ; PRIOR FILING DATE: 1997-10-29  
 ; PRIOR APPLICATION NUMBER: 60/063733  
 ; PRIOR FILING DATE: 1997-10-29  
 ; PRIOR APPLICATION NUMBER: 60/063735  
 ; PRIOR FILING DATE: 1997-10-29  
 ; PRIOR APPLICATION NUMBER: 60/063738  
 ; PRIOR FILING DATE: 1997-10-29  
 ; PRIOR APPLICATION NUMBER: 60/063755  
 ; PRIOR FILING DATE: 1997-10-17  
 ; PRIOR APPLICATION NUMBER: 60/064248  
 ; PRIOR FILING DATE: 1997-11-03  
 ; PRIOR APPLICATION NUMBER: 60/064809  
 ; PRIOR FILING DATE: 1997-11-07  
 ; PRIOR APPLICATION NUMBER: 60/065186  
 ; PRIOR FILING DATE: 1997-11-12  
 ; PRIOR APPLICATION NUMBER: 60/065846  
 ; PRIOR FILING DATE: 1997-11-17  
 ; PRIOR APPLICATION NUMBER: 60/066364  
 ; PRIOR FILING DATE: 1997-11-21  
 ; PRIOR APPLICATION NUMBER: 60/066453  
 ; PRIOR FILING DATE: 1997-11-24  
 ; PRIOR APPLICATION NUMBER: 60/066511  
 ; PRIOR FILING DATE: 1997-11-24  
 ; PRIOR APPLICATION NUMBER: 60/066770  
 ; PRIOR FILING DATE: 1997-11-24  
 ; PRIOR APPLICATION NUMBER: 60/069212  
 ; PRIOR FILING DATE: 1997-12-11  
 ; PRIOR APPLICATION NUMBER: 60/069278  
 ; PRIOR FILING DATE: 1997-12-11  
 ; PRIOR APPLICATION NUMBER: 60/069334  
 ; PRIOR FILING DATE: 1997-12-11  
 ; PRIOR APPLICATION NUMBER: 60/069694  
 ; PRIOR FILING DATE: 1997-12-16  
 ; PRIOR APPLICATION NUMBER: 60/072320  
 ; PRIOR FILING DATE: 1998-01-23  
 ; PRIOR APPLICATION NUMBER: 60/073612  
 ; PRIOR FILING DATE: 1998-02-04  
 ; PRIOR APPLICATION NUMBER: 60/074086  
 ; PRIOR FILING DATE: 1998-02-09  
 ; PRIOR APPLICATION NUMBER: 60/074092  
 ; PRIOR FILING DATE: 1998-02-09  
 ; PRIOR APPLICATION NUMBER: 60/077791  
 ; PRIOR FILING DATE: 1998-03-12  
 ; PRIOR APPLICATION NUMBER: 60/078910  
 ; PRIOR FILING DATE: 1998-03-20  
 ; PRIOR APPLICATION NUMBER: 60/079294  
 ; PRIOR FILING DATE: 1998-03-25  
 ; PRIOR APPLICATION NUMBER: 60/079663  
 ; PRIOR FILING DATE: 1998-02-27  
 ; PRIOR APPLICATION NUMBER: 60/079728  
 ; PRIOR FILING DATE: 1998-03-27  
 ; PRIOR APPLICATION NUMBER: 60/080165  
 ; PRIOR FILING DATE: 1998-03-31  
 ; PRIOR APPLICATION NUMBER: 60/081203  
 ; PRIOR FILING DATE: 1998-04-09  
 ; PRIOR APPLICATION NUMBER: 60/081229  
 ; PRIOR FILING DATE: 1998-04-09  
 ; PRIOR APPLICATION NUMBER: 60/081695  
 ; PRIOR FILING DATE: 1998-04-14  
 ; PRIOR APPLICATION NUMBER: 60/081817  
 ; PRIOR FILING DATE: 1998-04-15  
 ; PRIOR APPLICATION NUMBER: 60/081818  
 ; PRIOR FILING DATE: 1998-04-15  
 ; PRIOR APPLICATION NUMBER: 60/082999  
 ; PRIOR FILING DATE: 1998-04-24  
 ; PRIOR APPLICATION NUMBER: 60/083322  
 ; PRIOR FILING DATE: 1998-04-28  
 ; PRIOR APPLICATION NUMBER: 60/083545  
 ; PRIOR FILING DATE: 1998-04-29  
 ; PRIOR APPLICATION NUMBER: 60/084600  
 ; PRIOR FILING DATE: 1998-05-07  
 ; PRIOR APPLICATION NUMBER: 60/084627  
 ; PRIOR FILING DATE: 1998-05-07  
 ; PRIOR APPLICATION NUMBER: 60/084637  
 ; PRIOR FILING DATE: 1998-05-07  
 ; PRIOR APPLICATION NUMBER: 60/085149



; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P330R1C182  
 ; CURRENT APPLICATION NUMBER: US/10/140,808  
 ; CURRENT FILING DATE: 2002-05-07  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 135  
 ; LENGTH: 1675  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-140-808-135

Query Match 37.6%; Score 691.6; DB 5; Length 1675;  
 Best Local Similarity 84.5%; Pred. No. 7.1e-197;  
 Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

Qy	5	CGAGGAGGAGCGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	64
Db	287	CGCGGAGGAGGAGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	346
Qy	65	TGCTGAGCTCGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	124
Db	347	TGCTGAGCTCGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	406
Qy	125	TCCTTCCTGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	184
Db	407	TGCTGCGGCTCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	466
Qy	185	GATTGCCAGCAATGCGAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	244
Db	467	GCCTGCCAGCAATGCGAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	526
Qy	245	AGGAAACGGGAGGAGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	304
Db	527	AGGAAACGGGAGGAGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	586
Qy	305	GCTCTGAGTCAAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	364
Db	587	CCTCTGAGTCAAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	646
Qy	365	GCAAGAGGCTTCTGAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	424

Db 1127 TCCAGAGGATCCCTCTTACACAGAGCGCGGCTGGGCTGGGCGCATGAGTCCA 1186  
 Qy 899 TCCTTGAATCAAGACCCCTGCTGCTGCTGAGTGTGGGA 934  
 Db 1187 CCTCTGCTGAGAGCCCTGATTTTGAAGCTGAGGA 1222

RESULT 5

US-10-121-049-135  
 ; Sequence 135, Application US/10121049  
 ; Publication No. US20030022239A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P330R1C17  
 ; CURRENT APPLICATION NUMBER: US/10/121,049  
 ; CURRENT FILING DATE: 2002-04-12  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 135  
 ; LENGTH: 1675  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-121-049-135

Query Match 37.6%; Score 691.6; DB 5; Length 1675;  
 Best Local Similarity 84.5%; Pred. No. 7.1e-197;  
 Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

Qy	5	CGAGGAGGAGCGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	64
Db	287	CGCGGAGGAGGAGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	346
Qy	65	TGCTGAGCTCGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	124
Db	347	TGCTGAGCTCGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	406
Qy	125	TCCTTCCTGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	184
Db	407	TGCTGCGGCTCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	466
Qy	185	GATTGCCAGCAATGCGAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	244
Db	467	GCCTGCCAGCAATGCGAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	526
Qy	245	AGGAAACGGGAGGAGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	304
Db	527	AGGAAACGGGAGGAGCGCGGCTCGCTCTGCTGCGGCTGGGCTGGGCGCATGAGTCCA	586
Qy	305	GCTCTGAGTCAAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	364
Db	587	CCTCTGAGTCAAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	646
Qy	365	GCAAGAGGCTTCTGAGTGTGCAAGTGTGCTGCTGAGTGAAGTGGCTTTTG	424

Db 647 GCAAGAGCTCTCGATTATAGCTGTCACAAAGGAGAGGACCGCGCAGCAATCGATTTGGCA 706  
Qy 425 AGGGTATCTCGAGACCTTTGAGACGCTGCAACACTAGTCCACAAAGGGGTCAAGGTGG 484  
Db 707 AGGCATGTGAGAGACCTTTGAGACATTTACAACTGTGTATACAAAGGGGTCAAGGTGG 766  
Qy 485 TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTTCAGCAGAGGTGGCTGACCTCAAGA 544  
Db 767 TGATGGACATCCCTATGAGCTGTGGAAACGAGACTTCTGACAGAGTGGCTGACCTCAAGA 826  
Qy 545 AGCAGTGTGACGTGCTGTGGAAGATTGGAAGGTGATTCAGGACTGGTACAGGAACC 604  
Db 827 AGCAGTGTGATGTGCTGTGGAAGATTGGAAGGTGATTCAGGACTGGTACAGGAACC 886  
Qy 605 ACCAGGAGGAAGACTGACTGAATTTCTCTGTGCCAACACAGCTGTGAAGGGAAGGACA 664  
Db 887 ACCAGGAGGAAGACTGACTGAATTTCTCTGTGCCAACACAGCTGTGAAGGGAAGGACA 946  
Qy 665 CGAGTTGCTTAGCAGAGCGGTGGTCTGGCAAGAAAGGGGACATAGCCTCCCTGGGAGGGA 724  
Db 947 CCAGTTGCTTGCAGAGCAGTGGTCCGCGCAAGAGGGAGACACAGCTGCCCTGGGAGGGA 1006  
Qy 725 AGAATCCAAGAGAGCGAGCGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778  
Db 1007 AGAAGTCCAAGAGAGCAGCAGCGGCGCAAGGAGCAGCGGCGCAGGAGTAGCAGCAGCA 1066  
Qy 779 AGCAGGAAGAACTGGGGGCGCTGGGGAGGATGCCAACCGCAGGAGGAGGAGGTG 838  
Db 1067 AACAAAGGAAGAGCTGGGTGGCTTGGAGGAGACCCAGCCCGCAGGAGGATGAGGGA 1126  
Qy 839 TGCAGAGGCATCGCCCTCCACACACAGCCCTGATGAGCTCTGAGCCACCACGAGCAT 1186  
Db 1127 TCCAGAGGCATCCCTCTCACACAGCCCTGATGAGCTCTGAGCCACCACGAGCAT 1186

## RESULT 6

US-10-123-904-135  
; Sequence 135, Application US/10123904  
; Publication No. US20030022328A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C54  
; CURRENT FILING DATE: 2002-04-16  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 135  
; LENGTH: 1675  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-123-904-135

Query Match 37.6%; Score 691.6; DB 5; Length 1675;  
Best Local Similarity 84.5%; Pred. No. 7.1e-197;  
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;  
Qy 5 CGAGGAGGAAGCCGCCCGGGTCCGCTCTGCTCTGGGTCCGGCTAGAGTCCA 64  
Db 287 CGCGAGGAGGAACCCGCCCGGTCTTTTAGGGTCCGGCCCGGCCCATGGATTCAA 346  
Qy 65 TGTCTGAGCTCGCCCGCGCTGCTCTTATTTCTTTGCTGCTGCTGCTCGGTCTGC 124  
Db 347 TGCCTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCCCTGCTGCTGCTGCTGCTG 406  
Qy 125 TCCTTCTTCCCTCCGAAAGCTAGGCCCGAGTCCCGCCGGGCTGAGGAGACCGACTGGGTGC 184  
Db 407 TGCTGCCGCCCGCGAGCTGSGCCCGAGCCAGGCGCGAGCTGAGGAGAACGACTGGGTTC 466  
Qy 185 GATTGCCCAGCAATGGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTG 244  
Db 467 GCCTGCCCAGCAATGCAAGTGTGTAATATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG 526  
Qy 245 AGGAAACGGGAAGACCAAGGAAGTGTGACACCGGCTATGGCATCCTGACCGGAGG 304  
Db 527 AGGAAACCGGCAAGACCAAGGAGGTGATTGGCACGGGCTATGGCATCCTGACCCAGAGG 586  
Qy 305 GCTCTGAGTCAAGTACACCAAGTCGGACTTTACCGTTAAATTGAAGTCACTGAGACCATTT 364  
Db 587 CCTCTGGAGTCAANATACCAAGTCGGACTTTCGGTTAAATCGAAGTCACTGAGACCATTT 646  
Qy 365 GCAAGAGGCTTCTGGACTACAGCTGCAAGGAGAGAGACTGGCAGCAACCGGTTTGGCA 424  
Db 647 GCAAGAGGCTCTCGATTATAGCCTGCAAGGAGAGAGACCGGCAAGCAATCGATTTGGCA 706  
Qy 425 AGGTTATGTCGGAGACCTTTGAGACGCTGCAACACTAGTCCACAAAGGGGTCAAGGTGG 484  
Db 707 AGGCATGTGAGAGACCTTTGAGACATTAACAACCTTGGTTACAAAGGGGTCAAGGTGG 766  
Qy 485 TGATGGATATCCCTCTATGAGCTGTGGAAACGAGACTTCTGACAGAGTGGCTGACCTCAAGA 544  
Db 767 TGATGGACATCCCTCTATGAGCTGTGGAAACGAGACTTCTGACAGAGTGGCTGACCTCAAGA 826  
Qy 545 AGCAGTGTGAGCTGCTGGTGGAAAGATTGGAAGGTGATTGAGGACTGGTACAGGAACC 604  
Db 827 AGCAGTGTGATGTGCTGGTGGAAAGATTGAGGAGGTGATCGAGGACTGGTACAGGAACC 886  
Qy 605 ACCAGGAGGAAGACCTGACTGAATTTCTCTGTCACACCGCTGCTGAAGGGAAGGACA 664  
Db 887 ACCAGGAGGAAGACCTGACTGAATTTCTCTGTCGCGCAACCCAGCTGTGAAGGGAAGGACA 946  
Qy 665 CGAGTTGCTTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCCTCCCTGGGAGGGA 724  
Db 947 CCAGTTGCTTGGCAGAGCAGTGGTCCGCGCAAGAGGAGAGACACAGCTGCCCTGGGAGGGA 1006  
Qy 725 AGAATCCAAGAAAGCGCAGCGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778  
Db 1007 AGAAGTCCAAGAAAGAGCAGCAGCGGCGCAAGGAGCAGCGGCGGAGGAGTAGGAGGGA 1066  
Qy 779 AGCAGAGGAAGAACTGGGGGCGCTGGGGAGGATGCCAACCGCAGGAGGAGGAGGTG 838  
Db 1067 AACAAAGGAAGAGCTGGGTGGCTTGAAGGAGACCCCGCAGGAGGATGAGGGA 1126  
Qy 839 TGCAGAGGCATCCCTCCACACACAGCCCTGATGAGCTGTGAGCCACCACGAGCAT 898  
Db 1127 TCCAGAGGCATCCCTCTCACACAGCCCTGATGAGCTCTGAGCCACCACGAGCAT 1186  
Qy 899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934  
Db 1187 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 1222

## RESULT 7

US-10-140-470-135  
; Sequence 135, Application US/10140470



```

Db 347 TGCCTGAGCCGCGTCCGCTGTCTCTCTGCTTCTTCCCTGCTGCTGCTGCTGTC 406
Qy 125 TCCTTCTGCCCCGAAGCTAGCCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTGC 184
Db 407 TGCTGCGGCCCCGAGAGCTGGCGCCGAGCCAGGCGGAGCTGAGGAGAACACTGGGTTC 466
Qy 185 GATTGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTG 244
Db 467 GCCTGCCAGCAAAATGCGAAGTGTGTAATATGTTGCTGTGGAGCTGAAGTCAGCCTTG 526
Qy 245 AGGAAACGGGAAAGAACCAAGAAATGATGACCGGCTATGGCATCTGTCGACCGGAGG 304
Db 527 AGGAAACGGGAAAGAACCAAGAGGTGATTGGCAGGGCTATGGCATCTGTCGACGAGG 586
Qy 305 GCTCTGGAGTCAAGTACACCAAGTCGGACTTACCGTTAAATGAAAGTCACTGAGACCATTT 364
Db 587 CCTCTGGAGTCAAAATACACCAAGTCGGACTTACCGTTAAATGAAAGTCACTGAGACCATTT 646
Qy 365 GCAAGAGGCTTCTGGACTACAGCCTGCAACAGGAGAGGACTGGCAGCAACCGGTTTGCCA 424
Db 647 GCAAGAGGCTTCTGGATTATAGCTTGCAACAGGAGAGGACCGGCGACATCGATTGCCA 706
Qy 425 AGGATATGTCGGAGACCTTTGAGAGCTGTCACAACTTAGTCCACAAAGGGGTCAAGGTGG 484
Db 707 AGGATATGTCAGAGACCTTTGAGACATTTACACAACTTGTATACAAAGGGGTCAAGGTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGACGAGACTTCAGCAGAGGTGGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGACGAGACTTCTGACAGAGGTGGCTGACCTCAAGA 826
Qy 545 GCAAGAGGCTTCTGGACTACAGCCTGCAACAGGAGAGGACTGGCAGCAACCGGTTTGCCA 424
Db 647 GCAAGAGGCTTCTGGATTATAGCTTGCAACAGGAGAGGACCGGCGACATCGATTGCCA 706
Qy 425 AGGATATGTCGGAGACCTTTGAGAGCTGTCACAACTTAGTCCACAAAGGGGTCAAGGTGG 484
Db 707 AGGATATGTCAGAGACCTTTGAGACATTTACACAACTTGTATACAAAGGGGTCAAGGTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGACGAGACTTCAGCAGAGGTGGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGACGAGACTTCTGACAGAGGTGGCTGACCTCAAGA 826
Qy 545 AGCAGTGTGAGCTGTGCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACC 604
Db 827 AGCAGTGTGATGTGCTGTGGAAGAGTTTGAAGAGGTGATCGAGGACTGGTACAGGAACC 886
Qy 605 ACCAGGAGGAAGACTGTAATGAAATTCCTGTGTCACCAACCGTGTGTAAGGGAAGAGACA 664
Db 887 ACCAGGAGGAAGACTGTAATGAAATTCCTGTGTCACCAACCGTGTGTAAGGGAAGAGACA 946
Qy 665 CGAGTTGCTAGCAGAGCGGTGGTCTGCAAGAGGGGGACATAGCCTCCCTGGGAGGGA 724
Db 947 CCAGTTGCTGTGAGAGCAGTGGTCCGCAAGAGGGGAGACACAGCTGCCCTGGGAGGGA 1006
Qy 725 AGAATCAAGAAAGAGCGCAGCGAGTCAAG-----GGCTCTCTCCAGTGGCAGCAGCA 778
Db 1007 AGAAGTCCAGAGAGAGAGCAGCAGCGGCCAAGGAGAGGCGGAGGATGAGCAGCAGCA 1066
Qy 779 AGCAGAGGAAGAACTGGGGGGCTTGGGGAGGATGCCAACCGCCGAGAGAGAGGAGGTG 838
Db 1067 AACAAAGGAAGAGCTGGGTGGCTTGAAGGAGACCCAGCGCCCGAGGAGATGAGGGA 1126
Qy 839 TGCAGAGGCAATCGCCCTCCACACAGCCCGCTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db 1127 TCCAGAGGCAATCGCCCTCCACACAGCCCGCTGATGAGCTGTGAGCCCGAGCAGCAT 1186
Qy 899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 1187 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 1222

```

RESULT 9

```

US-10-176-918-135
; Sequence 135, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.

```

```

; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 135
; LENGTH: 1675
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-918-135

```

```

Query Match 37.6%; Score 691.6; DB 5; Length 1675;
Best Local Similarity 84.5%; Pred. No. 7.1e-197; Indels 6; Gaps 1;
Matches 791; Conservative 0; Mismatches 139;

```

```

Qy 5 CGAGGAGAGGAGCGCCCGGCTCCGCTCTGCTCTGGGTCCGGCTGGGGCCATGGAGTCCA 64
Db 287 CGCGGAGAGGAGAACCGCCCGGCTCTTTAGGGTCCGGCGCGGCCCATGATTCAA 346
Qy 65 TGTCTGAGCTCGCGCCCGGCTGCTCTTATTTCTTTGCTGCTGCTGCTGCTGCTGCTGCTG 124
Db 347 TGCTGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 406
Qy 125 TCCTTCTGCCCCGAAGCTAGGCCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTGC 184
Db 407 TGCTGCGGCGCCCGGAGCTGGGCGCGGAGCCAGGCGGAGCTGAGGAGAACACTGGGTTC 466
Qy 185 GATTGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTG 244
Db 467 GCCTGCCAGCAAAATGCGAAGTGTGTAATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG 526
Qy 245 AGGAAACGGGAAAGAACCAAGAAATGATGACCGGCTATGGCATCTGTCGACCGGAGG 304
Db 527 AGGAAACGGGAAAGAACCAAGAGGTGATTGGCAGGGCTATGGCATCTGTCGACGAGG 586
Qy 305 GCTCTGGAGTCAAGTACACCAAGTCGGACTTACCGTTAAATGAAAGTCACTGAGACCATTT 364
Db 587 CCTCTGGAGTCAAAATACCAAGTCCGACTTTCGGGTTAATCGAAGTCACTGAGACCATTT 646
Qy 365 GCAAGAGGCTTCTGGACTACAGCCTGCAACAGGAGAGGACTGGCAGCAACCGGTTTGCCA 424
Db 647 GCAAGAGGCTTCTGGATTATAGCTTGCAACAGGAGAGGACCGGCGACATCGATTGCCA 706
Qy 425 AGGATATGTCGGAGACCTTTGAGAGCTGTCACAACTTAGTCCACAAAGGGGTCAAGGTGG 484
Db 707 AGGATATGTCAGAGACCTTTGAGACATTTACACAACTTGTATACAAAGGGGTCAAGGTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGACGAGACTTCAGCAGAGGTGGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGACGAGACTTCTGACAGAGGTGGCTGACCTCAAGA 826
Qy 545 AGCAGTGTGAGCTGTGCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACC 604
Db 827 AGCAGTGTGATGTGCTGTGGAAGAGTTTGAAGAGGTGATCGAGGACTGGTACAGGAACC 886
Qy 605 ACCAGGAGGAAGACTGTAATGAAATTCCTGTGTCACCAACCGTGTGTAAGGGAAGAGACA 664
Db 887 ACCAGGAGGAAGACTGTAATGAAATTCCTGTGTCACCAACCGTGTGTAAGGGAAGAGACA 946
Qy 665 CGAGTTGCTAGCAGAGCGGTGGTCTGCAAGAGGGGGACATAGCCTCCCTGGGAGGGA 724
Db 947 CCAGTTGCTGTGAGAGCAGTGGTCCGCAAGAGGGGAGACACAGCTGCCCTGGGAGGGA 1006

```



```
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 135
; LENGTH: 1675
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-865-135

Query Match      37.6%; Score 691.6; DB 5; Length 1675;
Best Local Similarity 84.5%; Pred. No. 7.1e-197;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

Qy      5 CGAGGGAGGAAGCGCGCGGTCGCTCTGGTCCGGCTGAGGCTGGGCCATGAGTCCA 64
Db      287 CGCGGAGAGGAACCGCGCGGTCCTTAGGGTCGGGCCCGCGGCCATGATTCAA 346

Qy      65 TGTCTGAGCTCGCGCCCGCTGCCCTCTTATTTCTTGTCTGCTGCTCCGCTGCTGC 124
Db      347 TGCCTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCCCTTGTCTGCTGCTGCTGC 406

Qy      125 TCCTTCTGCCCCGAAAGCTAGGCCCGAGTCCGCGCGGGCTGAGGAGACCACTGGGTGC 184
Db      407 TGCTCGCGCGCCCGAGCTGGGCCCGAGCCGAGCCGAGCTGAGGAGAACACTGGGFTC 466

Qy      185 GATTGCCAGCAATGCAAGTGTGCAAGTATTTGCTGTGAGCTGAAGTCGGCTTTTG 244
Db      467 GCCTGCCAGCAATGCAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTTG 526

Qy      245 AGGAAACCGGAAAGCAACCAAGGAGTGTGACACCGGCTATGGCATCTTGACCGGGAAG 304
Db      527 AGGAAACCGGCAAGCAACCAAGGAGTGTGACACCGGCTATGGCATCTTGACCGAAG 586

Qy      305 GCTCTGAGTCAAGTACCAAGTCCGACTTACCGGCTTAAATTGAAGTCACTGAGACCATTT 364
Db      587 CCTCTGAGTCAAAATACCAAGTCCGACTTTCGGTTTAAATGAAAGTCACTGAGACCATTT 646

Qy      365 GCAGAGGCTTCTGAGCTACAGCTGCACAGGAGGAGCTGGCAGCAACCGTTTGCA 424
Db      647 GCAAGAGCTCTGTGATTTATAGCTTGCACAGGAGGAGCCGCGAGCAATCGATTTGCA 706

Qy      425 AGGGTATCTCGAGACCTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGG 484
Db      707 AGGGCATCTCAGAGACCTTTGAGACATTTACAACTGTGTACACAAAGGGTCAAGGTGG 766

Qy      485 TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGTGTGCTGACCTCAAGA 544
Db      767 TGATGGACATCCCTATGAGCTGTGGAAACGAGACTTCTGCAGAGTGTGCTGACCTCAAGA 826

Qy      545 AGCAGTGTGACGTGCTGCTGGAAGATTTGAGAGGTGATTTGAGACTGTGTACAGGAACC 604
Db      827 AGCAGTGTGATGCTGTGTGGAAGATTTGAGAGGTGATCGAGACTGTGTACAGGAACC 886

Qy      605 ACCAGAGGAAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAAGGGAAGGACA 664
Db      887 ACCAGAGGAAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAAGGGAAGGACA 946

Qy      665 CGAGTTGCTTACAGAGCGGTGCTGTCGAAGAAGAGGGGACATAGCCTCTCTGGAGGGA 724
Db      947 CCAGTTGCTTGCAGAGCAGTGGTCCGCGCAAGAGGGGAGACACAGCTGCCCTGGGAGGGA 1006

Qy      725 AGAATCCAGNAGAGCGCAGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778
Db      1007 AGAAGTCCAGNAGAGAGCAGCAGGCGCAAGGACGAGCGGCGAGGATGAGCAGCAGCA 1066

Qy      779 AGCAGAGGAAGAACTGGGGGCGCTTGGGGGAGGATGCCAAACCGCCGAGGAGGAGGGGTG 838
Db      1067 AACAAAGNAGAGAGTGGGTGGCCCTTGGAGGAGACCCAGCCCCGAGGAGGATGAGGGCA 1126

Qy      839 TGCAGAGGCATCGCCCTCCACACAGCCCCCTGTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db      1127 TCCAGAGGCATCCCTCTCTCACACAGCCCCCTGTGATGAGCTCTGAGCCCAACCCAGCAT 1186
```

```
Qy      899 TCCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGA 934
Db      1187 CCTCTGTCCTGAGACCCCTGATTTTGAAGCTGAGGA 1222
```

## RESULT 12

```
US-10-140-474-135
; Sequence 135, Application US/10140474
; Publication No. US20030032156A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C162
; CURRENT APPLICATION NUMBER: US/10/140,474
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 135
; LENGTH: 1675
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-474-135
```

```
Query Match      37.6%; Score 691.6; DB 5; Length 1675;
Best Local Similarity 84.5%; Pred. No. 7.1e-197;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;
```

```
Qy      5 CGAGGGAGGAAGCGCGCGGTCGCTCTGCTGCGGTCGGCTGGGCCATGAGTCCA 64
Db      287 CGCGGAGAGGAACCGCGCGGTCCTTAGGGTCGGGCCCGCGGCCATGATTCAA 346

Qy      65 TGTCTGAGCTCGCGCCCGCTGCCCTCTTATTTCTTGTCTGCTGCTCCGCTGCTGC 124
Db      347 TGCCTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCCCTTGTCTGCTGCTGCTGC 406

Qy      125 TCCTTCTGCCCCGAAAGCTAGGCCCGAGTCCGCGCGGGCTGAGGAGACCACTGGGTGC 184
Db      407 TGCTCGCGCGCCCGAGCTGGGCCCGAGCCGAGCCGAGCTGAGGAGAACACTGGGFTC 466

Qy      185 GATTGCCAGCAATGCAAGTGTGCAAGTATTTGCTGTGAGCTGAAGTCGGCTTTTG 244
Db      467 GCCTGCCAGCAATGCAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTTG 526

Qy      245 AGGAAACCGGAAAGCAACCAAGGAGTGTGACACCGGCTATGGCATCTTGACCGGGAAG 304
Db      527 AGGAAACCGGCAAGCAACCAAGGAGTGTGACACCGGCTATGGCATCTTGACCGAAG 586

Qy      305 GCTCTGAGTCAAGTACCAAGTCCGACTTACCGGCTTAAATTGAAGTCACTGAGACCATTT 364
Db      587 CCTCTGAGTCAAAATACCAAGTCCGACTTTCGGTTTAAATGAAAGTCACTGAGACCATTT 646

Qy      365 GCAGAGGCTTCTGAGCTACAGCTGCACAGGAGGAGCTGGCAGCAACCGTTTGCA 424
Db      647 GCAAGAGCTCTGTGATTTATAGCTTGCACAGGAGGAGCCGCGAGCAATCGATTTGCA 706
```



	Qy	425	AGG	TAT	GTC	GGA	GACCTTTTGAGACGCTGTGCACAACCTAGTCCACAAGGGGTCAAGTGG	484
	Db	707	AGG	CAT	GTG	CAG	AACCTTTTGAGACATTACAACTGGTACAAAGGGGTCAAGTGG	766
	Qy	485	TGA	TGA	TAT	CCCCTATGAGCTGTGGAAACGAGACTCAGCAGAGGTGGCTGACCTCAAGA	544	
	Db	767	TGA	TGG	AAT	CCCCCTATGAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA	826	
	Qy	545	AGC	AGT	GTG	AGCGTCTGGTGGAAAGAGTTTGAAGAGAGTCAATTGAGGACTGTGTACAGGAACC	604	
	Db	827	AGC	AGT	GTG	ATGTGCTGGTGGAAAGAGTTTGAAGAGAGTGAATCGAGGACTGGTACAGGAACC	886	
	Qy	605	ACC	AGG	AGG	AAGA	CCTGACTGAATTCCTCTGTGCCAACCAACGTCGTAAGAGGAAAAGACA	664
	Db	887	ACC	AGG	AGG	AAGA	CCTGACTGAATTCCTCTGCGCAACCAACGTCGTAAGAGGAAAAGACA	946
	Qy	665	CGA	GTT	GCCTTAGCAGAGCGGTGGTCTGGCAGAGAGGGGGACATAGCCTCCCTGGGAGGGA	724		
	Db	947	CCA	GTT	GCCTTGGCAGAGCAGTGGTCCCGCAAGNAGAGGAGACACAGCTGCCTCTGGGAGGGA	1006		
	Qy	725	AGA	AAT	CCA	AGA	AGCGCAGCGGAGTCAAG-----GGCTCTCAGTGGCAGCAGCA	778
	Db	1007	AGA	AGT	CCA	AGA	NAGACAGCAGAGGCCAAGCAGACAGGCGGCAGAGTAGCAGCAGCA	1066
	Qy	779	AGC	AGA	GGA	AGGA	ACTGGGGGGCCTGGGGGAGTAGTCCAAAGCCGAGGAGGAGAGGGGTG	838
	Db	1067	AAC	AA	GGA	AGG	AGACTGGTGGCTTTGAGGGAGACCCAGCCCGAGGAGGATGAGGGCA	1126
	Qy	839	TGC	AGA	AGG	CAT	CGCCCTCCCAACACAGCCCCCTGATGAGCTGTGAGCCGAGCTTAGTG	898
	Db	1127	TCC	AGA	AGG	CAT	CCCCCTCACACACAGCCCCCTGATGAGCTCTGAGCCCCCAGCAT	1186
	Qy	899	TC	CTT	GAT	CA	AGACCCCTGACTCAGACTTGGGA	934
	Db	1187	CCT	CTG	TCT	GAG	ACCCCTGATTTTGAAGCTGAGGA	1222

## RESULT 13

US-10-142-431-135  
; Sequence 135, Application US/10142431  
; Publication No. US20030036179A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: F3330R1C251  
; CURRENT APPLICATION NUMBER: US/10/142,431  
; CURRENT FILING DATE: 2002-05-10  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 135  
; LENGTH: 1675  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-142-431-135

### Query Match

37.6%; Score 691.6; DB 5; Length 1675;

## RESULT 14

US-10-143-114-135  
; Sequence 135, Application US/10143114  
; Publication No. US20030036180A1  
; GENERAL INFORMATION:



Qy	125	TCCTTCTGCCCCGAGCTAGCCGAGTCCGCGGGGCTGAGGAGACGACTGGGTGC	184
Db	407	TGCTGCGGCGCCCGAGCTGGGCGCGAGCCAGGCGGAGCTGAGGAGAACGACTGGGTTC	466
Qy	185	GATTGCCCGAGCAATCGAAGTGTGCAAGTATGTTGCTGTGTGAGCTGAAGTCGGCTTTTG	244
Db	467	GCCTGCCCGAGCAATCGAAGTGTGTAATATGTTGCTGTGTGAGCTGAAGTCAGCCTTTG	526
Qy	245	AGGAAACGGGAAAGACCAAGAAATGATGACACCGGCTATGGCATCTGGAACGGGAAGG	304
Db	527	AGGAAACGGGAAAGACCAAGGAGGTGATTGGCACGGGCTATGGCATCTGGACCAAGG	586
Qy	305	GCTCTGAGTCAAGTACACCAAGTCGGACTTTACGGTTAATTTGAAGTCACTGAGACCATTT	364
Db	587	CCTCTGAGTCAAAATACACCAAGTCGGACTTTGCGGTTAATCGAAGTCACTGAGACCATTT	646
Qy	365	GCAAGAGGCTTCTGGACTTACAGCCTGCAACAAGGAGAGGACTGGCAGCAACCGGTTTGCCA	424
Db	647	GCAAGAGGCTTCTGGATTATAGCTTACAGGAGGAGGACCGGCAAGCATCGATTTGCCA	706
Qy	425	AGGTAATGTCGAGACCTTTGAGACGCTTGCAACCTAGTCCACAAGGGGTCAAGGTGG	484
Db	707	AGGCAATGTCAGAGACCTTTGAGACATTTACACAACCTGGTACACAAGGGGTCAAGGTGG	766
Qy	485	TGATGGATATCCCTATAGCTGTGGACGAGACCTCAGCAGAGGTGGCTGACCTCAAGA	544
Db	767	TGATGGACATCCCTATAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA	826
Qy	545	AGCAGTGTGACGCTGCTGTGAAGATTGAAAGAGTGATTGAGGACTGGTACAGGAACC	604
Db	827	AGCAGTGTGCTGCTGTGTGAAGATTGAGGAGGTGATCGAGGACTGGTACAGGAACC	886
Qy	605	ACCAGGAGGAAGACCTGACTGAAATTCCTCTGTGCAACCAACGCTGTAAGGGAAAGGACA	664
Db	887	ACCAGGAGGAAGACCTGACTGAAATTCCTCTGTGCAACCAACGCTGTAAGGGAAAGGACA	946
Qy	665	CGAGTTGCTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCCTCCCTGGGAGGGA	724
Db	947	CCAGTTGCTGCGCAGAGCAGTGGTCCGGCAAGAGGGGAGACACAGCTGCCCCGGGAGGGA	1006
Qy	725	AGAAATCCAAGAAGCGCAGCGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA	778
Db	1007	AGAAATCCAAGAAGAGCAGCGAGTCAAGCGGCGGCGGCGGAGGAGTAGCAGCAGCA	1066
Qy	779	AGCAGAGGAAGAACTGGGGGCGCTGGGGGAGGATGCCAACCGCAGGAGGAGGAGGTG	838
Db	1067	AACAAGGAAGGAGCTGGGTGGCTTGAGGGGAGACCCCGAGGAGGATGAGGGA	1126
Qy	839	TGCAGAGGCAATCGCCCTCCACACAGCCCCCTGATGAGCTGTAGCCCGAGCTTAGTG	898
Db	1127	TCCAGAGGCAATCGCCCTCCACACAGCCCCCTGATGAGCTGTAGCCCGAGCTTAGTG	1186
Qy	899	TCCTTTGAATCAAGACCCCTGACTTCAGAGCTTGGA	934
Db	1187	CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA	1222

Search completed: December 30, 2005, 04:07:41  
Job time : 1386 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 23:12:51 ; Search time 931 Seconds  
(without alignments)  
1035.506 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacggaggaggaagcc.....taaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4172979 seqs, 262114271 residues

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq\*
- 2: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq\*
- 5: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq\*
- 7: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq2\*
- 9: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq3\*
- 10: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq\*

\*Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	691.6	37.6	1675	6	US-10-131-826A-135
2	104.4	5.7	946	6	US-10-750-185-37066
3	48.8	2.7	15376	7	US-11-121-086-5
4	42.6	2.3	168516	7	US-11-121-086-3
5	40.2	2.2	5468	6	US-10-821-234-49
6	40	2.2	1619	6	US-10-821-234-10
7	40	2.2	175416	7	US-11-121-086-43
8	39.2	2.1	191684	7	US-11-121-086-2
9	39	2.1	189539	7	US-11-121-086-16
10	37	2.0	1373	6	US-10-750-185-46147
11	37	2.0	1433	7	US-11-090-351-1
12	36.6	2.0	6691	7	US-11-186-284-3
13	36.6	2.0	1080000	6	US-10-928-446A-1
14	36.6	2.0	1080000	6	US-10-928-446A-181
15	36.6	2.0	1080000	6	US-10-928-446A-183
16	36.6	2.0	1080000	6	US-10-928-446A-185
17	36.6	2.0	1080000	6	US-10-928-446A-187
18	36.6	2.0	1080000	6	US-10-928-446A-189
19	36.6	2.0	1080000	6	US-10-928-446A-191
20	36.6	2.0	1080000	6	US-10-928-446A-193
21	36.6	2.0	1080000	6	US-10-928-446A-195
22	36.6	2.0	1080000	6	US-10-928-446A-197
23	36.6	2.0	1080000	6	US-10-928-446A-199

#### ALIGNMENTS

##### RESULT 1

US-10-131-826A-135

; Sequence 135, Application US/10131826A

; Publication No. US20050245730A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tamas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C128

; CURRENT APPLICATION NUMBER: US/10131,826A

; CURRENT FILING DATE: 2002-04-24

; PRIOR APPLICATION NUMBER: 60/049911

; PRIOR FILING DATE: 1997-06-18

; PRIOR APPLICATION NUMBER: 60/056974

; PRIOR FILING DATE: 1997-08-26

; PRIOR APPLICATION NUMBER: 60/059113

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059115

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059117

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059122

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059184

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/059263

; PRIOR FILING DATE: 1997-09-18

; PRIOR APPLICATION NUMBER: 60/059352

; PRIOR FILING DATE: 1997-09-19

; PRIOR APPLICATION NUMBER: 60/059588

Sequence 201, App  
Sequence 55, Appl  
Sequence 13415, A  
Sequence 13221, A  
Sequence 13233, A  
Sequence 8, Appl  
Sequence 52971, A  
Sequence 46060, A  
Sequence 455, App  
Sequence 453, App  
Sequence 57964, A  
Sequence 24, Appl  
Sequence 44, Appl  
Sequence 637, App  
Sequence 59, Appl  
Sequence 78059, A  
Sequence 51, Appl  
Sequence 1127, Ap  
Sequence 21, Appl  
Sequence 13424, A  
Sequence 13245, A  
Sequence 5006, Ap

C 24 36.6 2.0 1080000 6 US-10-928-446A-201  
25 36.4 2.0 2023 6 US-10-995-561-55  
26 36.4 2.0 12591 6 US-10-995-561-13415  
27 36.4 2.0 14023 6 US-10-995-561-13221  
28 36.4 2.0 199130 6 US-10-995-561-13233  
29 36.2 2.0 1191 7 US-11-061-869-8  
30 36 2.0 1062 6 US-10-750-185-52971  
31 36 2.0 2132 6 US-10-750-185-46060  
32 36 2.0 3156 6 US-10-995-561-455  
33 36 2.0 3201 6 US-10-995-561-453  
34 35.8 1.9 1128 6 US-10-750-185-57964  
35 35.8 1.9 171936 6 US-10-933-025-24  
36 35.4 1.9 167116 7 US-11-121-086-44  
37 35.2 1.9 1758 6 US-10-821-234-637  
38 35.2 1.9 6786 7 US-11-069-834-59  
39 35 1.9 201 6 US-10-995-561-78059  
40 35 1.9 7666 7 US-11-069-834-51  
41 34.8 1.9 600 6 US-10-750-185-1127  
42 34.8 1.9 10129 7 US-11-044-111-21  
43 34.8 1.9 16964 6 US-10-995-561-13424  
44 34.8 1.9 51749 6 US-10-995-561-13245  
45 34.4 1.9 201 6 US-10-995-561-5006

; PRIOR FILING DATE: 1997-09-19									
; Remaining Prior Application data removed - See File Wrapper or PALM.									
; NUMBER OF SEQ ID NOS: 550									
; SEQ ID NO 135									
; LENGTH: 1675									
; TYPE: DNA									
; ORGANISM: Homo Sapien									
US-10-131-826A-135									
Query Match 37.6%; Score 691.6; DB 6; Length 1675;									
Best Local Similarity 84.5%; Pred. No. 7.6e-169;									
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;									
QY	5	CGAGGGAGGAAGCGCCCGGGTCCGCTCTGCTCTCGGGTCCGGCTCGGGCCATGAGTCCA	64						
Db	287	CGCGGGAGGAGGAACCGCCCGGTCCTTTAGGGTCCGGGCCCGGCCCATGGAATCAA	346						
QY	65	TGTCCTGAGCTCGCGCCCGCTGCTCTTATTTCCTTTGCTGCTGCTCTCTCCGCTGCTGC	124						
Db	347	TGCTTGAGCCCGCTGCCGCTGCTCTCTGCTCTTTCCTTCTGCTGCTGCTGCTGCTGC	406						
QY	125	TCCTTCCTGCCCGAAGCTAGGCCCGGAGTCCCGCCGGGGCTGAGGAGACCGACTGGGTGC	184						
Db	407	TGCTGCCGGCCCGGAGCTGGGCCGAGCAGGCGCGAGCTGAGGAGAA CGACTGGGTTC	466						
QY	185	GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGCGCTTTTG	244						
Db	467	GCCTGCCCAGCAATGCGAAGTGTGTAATATATGTTGCTGTGGAGCTGAAGTCAAGCTTTG	526						
QY	245	AGGAAACGGGAAGACCAAGGAAGTGATGACACCGGTATGGCATCTCTGACGGGAGG	304						
Db	527	AGGAAACCGGACAGACCAGAGAGTGATTTGACCGGCTATGGCATCTCTGACAGAGG	586						
QY	305	GCTCTGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATTGAAGTCACTGAGACCAATTT	364						
Db	587	CCTCTGGAGTCAATATACACCAAGTCGGACTTTGCGGTTAAATCGAAGTCACTGAGACCAATTT	646						
QY	365	GCAAGAGGCTTCGAGCTACAGCCTGCACAAAGGAGAGACTTGGCAGCAACCGGTTTGCCA	424						
Db	647	GCAAGAGGCTCCTGGATTTATAGCCTGCACAAAGGAGAGGACCGGAGCAATCGAATTTGCCA	706						
QY	425	AGGGTATGTCGGAGACCTTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTGG	484						
Db	707	AGGGCATGTGAGAGACCTTTTGAGACATTAACACACTGTTACACAAAGGGGTCAAGGTGG	766						
QY	485	TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGAGGTGGCTGCACTCAAGA	544						
Db	767	TGATGGACATCCCCCTATGAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGCACTCAAGA	826						
QY	545	AGCAGTGTGAGTCTCGTGGAAAGAGTTTGAAGAGGTTGATTTGAGGACTGGTACAGGAACC	604						
Db	827	AGCAGTGTGATGTCTCGTGGAAAGAGTTTGAAGAGGTTGATCGAGGACTGGTACAGGAACC	886						
QY	605	ACCAGGAGGAAGACTGACTGAAATTCCTCTGTGCAACCAACCTGCTGAAGGGAAGGACA	664						
Db	887	ACCAGGAGGAAGACTGACTGAAATTCCTCTGCGCCACCACTGCTGAAAGGGAAGGACA	946						
QY	665	CGAGTTCCTAGCAGAGCGGTGCTGTGGCAAGAGGGGGACATAGCCTTCCTCGGAGGGA	724						
Db	947	CCAGTTCCTGCGAGAGCAGTGGTCCGGCAAGAAAGGGGAGACACAGCTGCCCTCGGAGGGA	1006						
QY	725	AGAAATCCAGAAAGCGGAGCGGAGTCAAG-----GGCTCTCTCCAGTGGCAGCAGCA	778						
Db	1007	AGAAATCCAGAAAGCAGCAGGAGGCAAGGCGAGCGGCGGAGGATGAGCAGCAGCA	1066						
QY	779	AGCAGAGGAAGGAATCTGGGGGGCTTGGGGGAGGATGCCAAACGCCGAGGAGGAGGAGGTG	838						
Db	1067	AACAAAGGAGGAGCTGGGTGGCCTTTGAGGGAGAGCCCGAGCCCGAGGAGGATGAGGGA	1126						
QY	839	TGCAGAAAGCAATCGCCCTCCCAACACAGCCCTCTGATGAGCTGTGAGCCCAAGCTTAGTG	898						
Db	1127	TCCAGAAAGGATCCCTCTTCAACACAGCCCTCTGATGAGCTGTGAGCCCAAGCAGCAT	1186						

	Query Match	2.3%	Score 42.6;	DB 7;	Length 168516;
	Best Local Similarity	47.5%;	Pred. No. 1.2;		
	Matches 160;	Conservative 0;	Mismatches 174;	Indels 3;	Gaps 1;
Qy	514	GAGACCTCAGAGGTGGCTGACCTCAAGAACCACTGTCAGTCTGTTGAAAGATTT	573		
Db	68433	GTGCCTTAACTTGAGGAGGAGGAGGGGGAGAGAGAGGAGGAGGGGAGGAAGCG	68492		
Qy	574	GAGAGGTGATTGAGNCTGGTACAGAACCAACGAGGAGAGACCTGACTCAATTCCTC	633		
Db	68493	GCAGAGAGCTTCGAGGCGAGGAGGAGGGGAGGAGGGGGGAGGAGTAAAGAGGTGGAA	68552		

```

RESULT 6
US-10-821-234-10/c
; Sequence 10, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A

```

```
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 10
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-10

Query Match          2.2%; Score 40; DB 6; Length 1619;
Best Local Similarity 49.1%; Pred. No. 0.98;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 639 CAACACGTGTGAAGGAAAGGACACGAGTTGCTTAGCAGAGCGGTGCTTGCAGAGAA 698
Db 1418 CAGCAAGAAGAAGAGGGCTACAATGTGAACGACGAGAAAGCCAAAGGAGAAACAAGAA 1359

Qy 699 GGGGGACATACCTCCCTCGGAGGGAGAAATCCAGNAGAGCGCAGCGGAGTCNAGGG 758
Db 1358 GGCCGAGGGCGCGACGAGGAGAGGGGACCCCGAAGGAGAGTGAGCCCCAGGCGCC 1299

Qy 759 CTCCTCCAGTGGCAGCAGCAAGCAGAGAGAACTGGGGGGCCTGGGGGAGGATGCCAA 818
Db 1298 CGCAGAGCCCGCGAGGCGCAAGGAGGGCAAGGAGAGCGCCACCAGGACCGGAGGCAA 1239

Qy 819 CGCCGAGGAGGAGGAGGTGTGCAGAGGCATCGCC 854
Db 1238 GGCCGAGGAGGAGGAGGCGAGAGGACGCGCGCGC 1203

RESULT 7
US-11-121-086-43
; Sequence 43, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 43
; LENGTH: 175416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-43

Query Match          2.2%; Score 40; DB 7; Length 175416;
Best Local Similarity 47.0%; Pred. No. 5.7;
Matches 156; Conservative 0; Mismatches 175; Indels 1; Gaps 1;

Qy 674 TAGCAGACGGTGTGCTGGCAAGAGGGGACATAGCCCTCCCTGGAGGGAGAAATCCA 733
Db 60882 TGGAAATAGTGTGGGGAGGAATAAATCGGAAGAAGAACGTGGGGAGAGATGGACGCTA 60941

Qy 734 AGAAGACGCGAGCGGAGTCAAGGCTCTCCAGTGGCA-CGACNAGCAGAGNAGAA 792
Db 60942 AGCAGAGAGATGGGGCGAGGAGAGGGGAACACAGCCACGCAATGGAAGAGCAAGAA 61001

Qy 793 CTGGGGGCGCTGGGGGAGGATGCCAAACCGCAGGAGGAGGAGGTGTGCAGAAAGGCATCG 852
Db 61002 GAGCAGGAGAGCGTGAAGGCGTGAGAACCGGAGCGCAGGCGAGGAGCTTCAGAGGCAGAA 61061

Qy 853 CCCCTCCACACAGCCCCCTGTAGTGTGAGCCCCAGCTAGTGTCTTGAATCAAGA 912
Db 61062 GCGCTTCAGAGCTGTGCAAAATCCAAATGCTAAGAAAGAACATGCAGTTCATTTGACRAGG 61121
```

---

```
Qy 913 CCCCTGACTTCAGAGCTTGGGACACGCAAGCGCAGCGCAGCGCAGCTCCAGCAAGGACA 972
Db 61122 GCAGGTCAGAAATGCCACCGTGCACAGAGGATGGAAGGCCAGGACCGCGCCCAAG 61181

Qy 973 GCTGCTGTCCAGCATCAGGCTCTCCCTCTGG 1004
Db 61182 CAGGGCTGTGGGAGCTGGTCCCTCTCTGG 61213

RESULT 8
US-11-121-086-2
; Sequence 2, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 191684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-2

Query Match          2.1%; Score 39.2; DB 7; Length 191684;
Best Local Similarity 43.2%; Pred. No. 9.5;
Matches 185; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

Qy 425 AGGATATGTCGGAGACCTTTGAGACGCTGCACACCTAGTCCACAAAGGGGTCAAGGTGG 484
Db 50266 AGGAGGTGGCGGAGGTAGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGG 50325

Qy 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544
Db 50326 TGGAGGTGAAGGTGGAGGAGGTAGAGGTAGAGGTGGAGGTGGAGGTAAATGGAGGTGG 50385

Qy 545 AGCAGTGTGACGTCTGCTGGAAGAGTTTGAAGAGGTGATTTAGGACTGTTACAGGAACC 604
Db 50386 AAGTGGAGGTGGAGGTGATGTTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGG 50445

Qy 605 ACCAGAGGAAAGACCTGACTGAATTCCTCTGTGTCCAAACCAAGCTCTGAAGGAAAGACA 664
Db 50446 TGGAGGTGGTGGAGTTGCTCGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGG 50505

Qy 665 CGAGTTCCCTAGCAGAGCGGTGCTTGGCAAGAGGGGACATAGCCCTCCCTGGAGGGA 724
Db 50506 AGGAGATGCTGGAGGTGGAGGTGATGGAGATGGAGGTGGAGGTGGAGGTGGAGGTGG 50565

Qy 725 AGAAATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAGCAGA 784
Db 50566 TGAAGGAAGAGGTCTCTGGAGGTGGAGGTGATGGAGGTGGAGGTGGAGGTGGAGGTGG 50625

Qy 785 GGAAGGAACCTGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGAGGTGTGCAGA 844
Db 50626 TGAAGGAGGAGGTGGAGGTGCTGGAGGTGGTGGAGGAGGAGGAGGTGCTGGAGTTGGAGG 50685

Qy 845 AGGCATCG 852
Db 50686 TGATATGG 50693

RESULT 9
US-11-121-086-16
; Sequence 16, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
```



[illegible]

```

; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPW01-029F2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR FILING DATE: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 6691
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82)...(3717)
; US-11-186-284-3

Query Match      2.0%; Score 36.6; DB 7; Length 6691;
Best Local Similarity 58.9%; Pred. No. 13;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 52 GCCATGGAGTCATGCTGTAGCTGCGCGCGCTGCTTTATTTCTTTGCTGCTG 111
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79 GCCATGGATCCGCGCGGAGCGCTGCGCGCTGCTGCTGCGCGCTGCTGCTG 138

Qy 112 CTTCCGCTGCTGCTTCTGTCGCGCGAGCTAGCGCGAGTCCGCG 158
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139 CTGCTGCTGCTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCG 185

RESULT 13
US-10-928-446A-1/c
; Sequence 1, Application US/10928446A
; Publication No. US2005027123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
; TITLE OF INVENTION: VIRAL BUDDING
; FILE REFERENCE: 0274-5785.1US
; CURRENT APPLICATION NUMBER: US/10/928,446A
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825234)
; OTHER INFORMATION: the 'n' at position 825234 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825270)
; OTHER INFORMATION: the 'n' at position 825270 may be 'c' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825401)
; OTHER INFORMATION: the 'n' at position 825401 may be 'c' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825428)
; OTHER INFORMATION: the 'n' at position 825428 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825473)
; OTHER INFORMATION: the 'n' at position 825473 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825765)
; OTHER INFORMATION: a "c" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825828)
; OTHER INFORMATION: the 'n' at position 825828 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826041)
; OTHER INFORMATION: the 'n' at position 826041 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826546)
; OTHER INFORMATION: the 'n' at position 826546 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826654)
; OTHER INFORMATION: the 'n' at position 826654 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826826)
; OTHER INFORMATION: the 'n' at position 826826 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826863)
; OTHER INFORMATION: the 'n' at position 826863 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (827008)
; OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (843055)
; OTHER INFORMATION: the 'n' at position 843055 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (843118)
; OTHER INFORMATION: the 'n' at position 843118 may be 't' or 'c'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (871027)
; OTHER INFORMATION: the 'n' at position 871027 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (871140)
; OTHER INFORMATION: the 'n' at position 871140 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (871168)
; OTHER INFORMATION: the 'n' at position 871168 may be 'c' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (872678)
; OTHER INFORMATION: the 'n' at position 872678 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (872742)
; OTHER INFORMATION: the 'n' at position 872742 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (925859)
; OTHER INFORMATION: the 'n' at position 925859 may be 'c' or 'g'
```





```

; NAME/KEY: CDS
; LOCATION: (993104)..(993154)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (999547)..(999608)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1000354)..(1000456)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1002118)..(1002284)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1006117)..(1006249)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1007860)..(1008036)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1010940)..(1011014)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1018160)..(1018291)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1018800)..(1018919)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1020028)..(1020225)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1026659)..(1026736)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1028113)..(1028167)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1034316)..(1034374)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1041390)..(1041455)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1043121)..(1043350)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1044868)..(1044989)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1047519)..(1047589)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1050296)..(1050391)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: (1060368)..(1060441)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1062648)..(1062708)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1064561)..(1064620)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1066207)..(1066314)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1067768)..(1067864)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1068609)..(1068681)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1073289)..(1073388)
; OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-183

Query Match      2.0%; Score 36.6; DB 6; Length 1080000;
Best Local Similarity 50.9%; Pred. No. 78;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 1547 CTGGGTCTCATCAGCCTCCAGAGTGTCCCTCATCGATCTTTTGTGCTTGTCCCA 1606
Db 236820 CTCGGCTCACTGCAACCTCCACCTCATGGGCTCAAGCAGTTCTTTGCTCAGCCTCCA 236761

Qy 1607 ATCCAGGGGCTGGAAGGCCATCACCATCATTTGAGGCTTAACCTGTCTAGTAGGAG 1666
Db 236760 AGTAGCTGGGATTACAGGCACCCACCACCATGCTGGCTAATTTTGTATTTTAGTAG 236701

Qy 1667 GTGCTGGGAGCGCCGGGTTGTTGGGTAATCACTCACTGGCTCTCAG 1717
Db 236700 AGACTGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTGACCTTAG 236650

```

Search completed: December 30, 2005, 04:23:29  
Job time : 941 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 03:38:26 ; Search time 9079 Seconds  
(without alignments)  
11513.945 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacgagggaagcgc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 5883141 seqs, 28421725653 residues

Word size : 0

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_in.\*  
3: gb\_env.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pr.\*  
9: gb\_ro.\*  
10: gb\_sts.\*  
11: gb\_sy.\*  
12: gb\_un.\*  
13: gb\_vi.\*  
14: gb\_htg.\*  
15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1780	96.8	1882 9 BC013549	BC013549 Mus muscu
2	1477	80.3	1909 9 AF361644	AF361644 Mus muscu
3	577	31.4	1655 9 AK220209	AK220209 Mus muscu
4	125	6.8	10892 14 AC117716	AC117716 Mus muscu
5	99	5.4	89855 14 AC074220	AC074220 Mus muscu
6	69	3.8	22256 14 AC134482	AC134482 Rattus no
7	69	3.8	246404 14 AC131887	AC131887 Rattus no
8	64	3.5	270958 14 AC135655	AC135655 Rattus no
9	50	2.7	526 6 CQ921127	CQ921127 Sequence
10	50	2.7	580 6 AX198865	AX198865 Sequence
11	50	2.7	580 6 AX209392	AX209392 Sequence
12	50	2.7	612 8 AF161347	AF161347 Homo sapi
13	50	2.7	904 8 BC022093	BC022093 Homo sapi
14	50	2.7	1201 6 AR379595	AR379595 Sequence
15	50	2.7	1277 8 BC008133	BC008133 Homo sapi
16	50	2.7	1413 8 BC008898	BC008898 Homo sapi
17	50	2.7	1413 8 BC008961	BC008961 Homo sapi
18	50	2.7	1593 8 BC011767	BC011767 Homo sapi

19	50	2.7	1675	6	AX464002	Sequence
20	50	2.7	1675	8	AY358960	Homo sapi
21	50	2.7	1720	8	BC004423	Homo sapi
22	50	2.7	4469	8	AK090425	Homo sapi
23	50	2.7	113109	8	HS475N16	Human DNA
24	49	2.7	2623	8	AK124196	Homo sapi
25	45	2.4	958	6	CQ716533	Sequence
26	45	2.4	962	8	HS080744	Homo sapien
27	36	2.0	1534	6	AR379729	Sequence
28	35	1.9	632	10	BV059085	SV212P6851
29	35	1.9	1657	6	BD194905	86 human
30	35	1.9	1657	6	CQ855233	Sequence
31	33	1.8	1658	6	BD218541	71 human
32	30	1.6	4881	9	AY267461	Mus muscu
33	30	1.6	165733	14	AC107706	AC107706 Mus muscu
34	30	1.6	187371	14	AC164172	AC164172 Mus muscu
35	30	1.6	212943	9	AC107787	AC107787 Mus muscu
36	30	1.6	213550	9	AC153370	AC153370 Mus muscu
37	29	1.6	103129	9	AL662809	Mouse DNA
38	29	1.6	171820	9	AC126550	Mus muscu
39	29	1.6	200057	4	CR956379	Pig DNA 8
40	29	1.6	216315	9	AL928963	Mouse DNA
41	29	1.6	258622	14	AC020877	AC020877 Mus muscu
42	28	1.5	1549	5	BC097110	Danio rer
43	28	1.5	114568	14	AC134480	Rattus no
44	28	1.5	148679	9	AC138308	Mus muscu
45	28	1.5	166302	5	CR385039	Zebrafish

#### ALIGNMENTS

RESULT 1

BC013549

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BC013549 1882 bp mRNA linear ROD 28-SEP-2004  
Mus musculus trinucleotide repeat containing 5, mRNA (cdna clone)  
MGC:19340 IMAGE:4222133, complete cds.

BC013549  
BC013549  
MGC.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1882)  
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Sheezy, T.E., Brownstein, M.J., Usdin, T.B., Toohy, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullen, S.J., Bosak, S.A., McKwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smal, D.E.,  
Schnerch, A., Schein, J.B., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

2 (bases 1 to 1882)  
Director MGC Project.  
Direct Submission  
Submitted (04-SEP-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer

REMARK  
COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 25 Row: J Column: 9  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31982623.

FEATURES  
source

Location/Qualifiers  
1..1882  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MGC:19340 IMAGE:4222133"  
/tissue\_type="Kidney, normal. 5 month old male mouse."  
/clone\_lib="NCI\_COAP\_Kid14"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

gene

1..1882  
/gene="Tnrc5"  
/note="synonyms: CAG4A, ERDA5"  
/db\_xref="GeneID:72029"  
/db\_xref="MGI:1919279"  
91..921  
/gene="Tnrc5"

CDS

/codon\_start=1  
/product="crinucleotide repeat containing 5"  
/protein\_id="AAH13549.1"  
/db\_xref="GI:15488836"  
/db\_xref="GeneID:72029"  
/db\_xref="MGI:1919279"  
/translation="MESMSLAPRCLLPPLLLPLLLPAPKLGPSPAGAEETDWR  
LPSKEVCYVAVELKAFETGKEVIDTGILDKGSGVKFTSDLRLEIVET  
ICKLLDYSLHKEGTGNRFAGMSSETFLHNLVHKGVKVMIDIPYELMNETSAEVA  
DLKKQCDYLVEFEVIEDWYRNHQEEDLTFELCANHYLKGKDTSCLAERWSGKKGDI  
ASLGGKSKKRSVKSGSSGSKQKELGLGLEDANAEEBEGVQKASPLPHSPDDEL  
"

ORIGIN

Query Match 96.8%; Score 1780; DB 9; Length 1882;  
Best Local Similarity 99.9%; Pred. NO. 0;  
Matches 1830; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 9 GGAGGAAGCGCGCGGTCGCTGCTGCTGCGGTCGGGTCGGGCATGGAGTCCATGTC 68  
Db 45 GGAGGAAGCGCGCGGTCGCTGCTGCTGCGGTCGGGTCGGGCATGGAGTCCATGTC 104

Qy 69 TGAGCTCGCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 128  
Db 105 TGAGCTCGCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 164

Qy 129 TCCTGCCCGGAGCTAGGCGCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGCGATT 188  
Db 165 TCCTGCCCGGAGCTAGGCGCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGCGATT 224

Qy 189 GCCAGCAAAATGCGAAGTGTGAAGTATGTTGCTGTGAGTGAAGTGTGAGGATTTGAGGA 248  
Db 225 GCCAGCAAAATGCGAAGTGTGAAGTATGTTGCTGTGAGTGAAGTGTGAGGATTTGAGGA 284

Qy 249 AACGGGAAGACCAAGGAAGTATTGACACCGGCTATGGCATCTCTGGACCGGAAGGGCTC 308

Db 285 AACGGGAAGACCAAGGAAGTATTGACACCGGCTATGGCATCTCTGGACGGGAAGGGCTC 344  
Qy 309 TGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATGAAGTCACTGAGACCAATTTGCAA 368  
Db 345 TGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATGAAGTCACTGAGACCAATTTGCAA 404  
Qy 369 GAGGCTTCTGAGCTACAGCCTGCAACAAGGAGAGACTGGCAGCAACCGGTTTGCCAAAGG 428  
Db 405 GAGGCTTCTGAGCTACAGCCTGCAACAAGGAGAGACTGGCAGCAACCGGTTTGCCAAAGG 464  
Qy 429 TATGTCGAGACCTTTGAGACGCTGCAACAACCTAGTTCACAAGAGGGTCAAGGTGGTAT 488  
Db 465 TATGTCGAGACCTTTGAGACGCTGCAACAACCTAGTTCACAAGAGGGTCAAGGTGGTAT 524

Qy 489 GGATATCCCTATGAGCTGTGGAAGGAGACTCTGACGAGAGTGGTGGCTCAAGAAGCA 548  
Db 525 GGATATCCCTATGAGCTGTGGAAGGAGACTCTGACGAGAGTGGTGGCTCAAGAAGCA 584  
Qy 549 GTGTGAGCTGCTGCTGGAAGAGTTTGAAGAGTGAATGAGGACTGGTACAGGAACCA 608  
Db 585 GTGTGAGCTGCTGCTGGAAGAGTTTGAAGAGTGAATGAGGACTGGTACAGGAACCA 644

Qy 609 GGAGGAAGACCTGACTGAATTCCTCTGTGTCACAACCTGTCTGAAGGAAAGGACACGAG 668  
Db 645 GGAGGAAGACCTGACTGAATTCCTCTGTGTCACAACCTGTCTGAAGGAAAGGACACGAG 704

Qy 669 TTGCTAGCAGAGCGGTGGTCTGCAAGAGAGGGGACATAGCCTCCCTGGGAGGGAAGAA 728  
Db 705 TTGCTAGCAGAGCGGTGGTCTGCAAGAGAGGGGACATAGCCTCCCTGGGAGGGAAGAA 764

Qy 729 ATCCAAGAAGACGCGAGCGAGTCAAGGGCTCTCTCAAGTGGCAGCAGCAGCAGAGGAA 788  
Db 765 ATCCAAGAAGACGCGAGCGAGTCAAGGGCTCTCTCAAGTGGCAGCAGCAGCAGAGGAA 824

Qy 789 GGAATCTGGGGGCTCTGGGGAGGATGCCAACCGCAGAGAGAGAGGGTGTGCAGAAAGC 848  
Db 825 GGAATCTGGGGGCTCTGGGGAGGATGCCAACCGCAGAGAGAGAGGGTGTGCAGAAAGC 884

Qy 849 ATCGCCCTCCACACAGCCCCCTGATGAGTGTGAGCCAGCTTAGTGTCTTGAATC 908  
Db 885 ATCGCCCTCCACACAGCCCCCTGATGAGTGTGAGCCAGCTTAGTGTCTTGAATC 944

Qy 909 AAGACCCCTGACTTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCGCAGCTCCAGCAAG 968  
Db 945 AAGACCCCTGACTTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCTCCAGCAAG 1004

Qy 969 GACAGTGTGTCAGCATCAGGTCCTCTCTCTGGGTGTGCCCCCTTTCCTTCCCTTGA 1028  
Db 1005 GACAGTGTGTCAGCATCAGGTCCTCTCTCTGGGTGTGCCCCCTTTCCTTCCCTTGA 1064

Qy 1029 CAACAGCAGAGGTGGAGGATCTGGGGTGTGGGAGAGCGCACCCCAAGGGAAGAGGA 1088  
Db 1065 CAACAGCAGAGGTGGAGGATCTGGGGTGTGGGAGAGCGCACCCCAAGGGAAGAGGA 1124

Qy 1089 GGAGGAGCAGAAAGCAGCTCTCTTTTACACAGTCCCCCTCACGAGTCCCGGGTCCACC 1148  
Db 1125 GGAGGAGCAGAAAGCAGCTCTCTTTTACACAGTCCCCCTCACGAGTCCCGGGTCCACC 1184

Qy 1149 CAGCATCCCCAGGCTGAGATCCAGGCTCTCTCATGGAAGCTGAAGAGCATGAGGCAAT 1208  
Db 1185 CAGCATCCCCAGGCTGAGATCCAGGCTCTCTCATGGAAGCTGAAGAGCATGAGGCAAT 1244

Qy 1209 AAGATGCTACACAGGCCCCCTTACGACGAGAGGACTCCGTGAGCCTCAGCAGCCAGG 1268  
Db 1245 AAGATGCTACACAGGCCCCCTTACGACGAGAGGACTCCGTGAGCCTCAGCAGCCAGG 1304

Qy 1269 CCTGCTCTTCTTCCCAAGCAATTCCTTCTGCTGGTCTCTGTCGATGCTGTAATTCG 1328  
Db 1305 CCTGCTCTTCTTCCCAAGCAATTCCTTCTGCTGGTCTCTGTCGATGCTGTAATTCG 1364

Qy 1329 AGAATCTTCAGGACAAACTCGGGTGTGCGACAAAGGGGCTGGACCCAGAGCCAGAGCCA 1388







Qy 757 GGCTCTCTCAGTGGCAGCAGCAGCAGGAGGAAGCACTGGGGGCTGGGGGAGGATGCC 816  
 Db 601 GGCTCTCTCAGTGGCAGCAGCAGCAGGAGGAAGCACTGGGGGCTGGGGGAGGATGCC 660  
 Qy 817 AACCCGAGGAGGAGGAGG 835  
 Db 661 AACCCGAGGAGGAGGAGG 679

RESULT 4  
 AC117716  
 LOCUS AC117716 108992 bp DNA linear HTG 20-DEC-2002  
 DEFINITION Mus musculus clone RP23-96H4, LOW-PASS SEQUENCE SAMPLING.  
 AC117716  
 VERSION AC117716.2 GI:27311376  
 KEYWORDS HTG; HTGS PHASE0  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.  
 1 (bases 1 to 108992)  
 Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP23-96H4

REFERENCE  
 AUTHORS  
 JOURNAL  
 TITLE  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 108992)  
 2 (bases 1 to 108992)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Baetien,V., Bloom,T., Boguslavskiy,L.,  
 Boukhaltier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,  
 Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 108992)  
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Baetien,V., Bloom,T., Boguslavskiy,L., Boukhaltier,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fero,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., MacLean,C., Macdonald,P., Major,J.,  
 Lindblad-Toh,K., Liu,G., Meneus,L., Meneus,L., Mienga,V.,  
 Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,  
 Rise,C., Rogov,P., Roman,J., Roy,A., Santos,R., Schauer,S., Schuback,R.,  
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,  
 Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (20-DEC-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Dec 20, 2002 this sequence version replaced gi:20128494.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L24204  
 Center clone name: 96\_H\_4

-----  
 \* NOTE: This record contains 88 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 \* 1141: contig of 1141 bp in length  
 \* 1142: gap of 100 bp  
 \* 1242: contig of 1137 bp in length  
 \* 2378: gap of 100 bp  
 \* 2379: contig of 1088 bp in length  
 \* 2479: gap of 100 bp  
 \* 3567: contig of 100 bp  
 \* 3667: gap of 100 bp  
 \* 4807: contig of 1140 bp in length  
 \* 4906: gap of 100 bp  
 \* 4907: contig of 1138 bp in length  
 \* 6044: gap of 100 bp  
 \* 6144: gap of 100 bp  
 \* 6145: contig of 1145 bp in length  
 \* 7289: gap of 100 bp  
 \* 7390: contig of 1131 bp in length  
 \* 8520: gap of 100 bp  
 \* 8521: contig of 1144 bp in length  
 \* 9764: gap of 100 bp  
 \* 9765: contig of 1115 bp in length  
 \* 10979: gap of 100 bp  
 \* 11079: contig of 1140 bp in length  
 \* 12219: gap of 100 bp  
 \* 12319: gap of 100 bp  
 \* 13466: contig of 1147 bp in length  
 \* 13467: gap of 100 bp  
 \* 13567: contig of 1104 bp in length  
 \* 14671: gap of 100 bp  
 \* 14770: contig of 1091 bp in length  
 \* 15861: gap of 100 bp  
 \* 15962: contig of 1153 bp in length  
 \* 17114: gap of 100 bp  
 \* 17214: contig of 1175 bp in length  
 \* 18389: gap of 100 bp  
 \* 18390: contig of 1165 bp in length  
 \* 18490: gap of 100 bp  
 \* 19655: contig of 1168 bp in length  
 \* 19754: gap of 100 bp  
 \* 20922: contig of 1155 bp in length  
 \* 21023: gap of 100 bp  
 \* 21277: contig of 1135 bp in length  
 \* 22178: gap of 100 bp  
 \* 22718: contig of 1135 bp in length  
 \* 23412: gap of 100 bp  
 \* 23512: contig of 1135 bp in length  
 \* 24647: gap of 100 bp  
 \* 24748: contig of 1091 bp in length  
 \* 25838: gap of 100 bp  
 \* 25938: contig of 1147 bp in length  
 \* 27085: gap of 100 bp  
 \* 27185: contig of 1147 bp in length  
 \* 28332: gap of 100 bp  
 \* 28432: contig of 1102 bp in length  
 \* 28433: gap of 100 bp  
 \* 29535: contig of 100 bp

```
*
* 29635 30724: contig of 1090 bp in length
* 30725 30824: gap of 100 bp
* 30825 33991: contig of 1167 bp in length
* 31992 32091: gap of 100 bp
* 32092 33257: contig of 1166 bp in length
* 33258 33357: gap of 100 bp
* 33358 34530: contig of 1173 bp in length
* 34531 34630: gap of 100 bp
* 34631 35794: contig of 1164 bp in length
* 35795 35894: gap of 100 bp
* 35895 37060: contig of 1166 bp in length
* 37061 37160: gap of 100 bp
* 37161 38311: contig of 1151 bp in length
* 38312 38411: gap of 100 bp
* 38412 39569: contig of 1158 bp in length
* 39570 39669: gap of 100 bp
* 39670 40816: contig of 1147 bp in length
* 40817 40916: gap of 100 bp
* 40917 42058: contig of 1142 bp in length
* 42059 42158: gap of 100 bp
* 42159 43278: contig of 1120 bp in length
* 43279 43378: gap of 100 bp
* 43379 44477: contig of 1099 bp in length
* 44478 44577: gap of 100 bp
* 44578 45746: contig of 1169 bp in length
* 45747 45846: gap of 100 bp
* 45847 47007: contig of 1161 bp in length
* 47008 47107: gap of 100 bp
* 47108 48266: contig of 1159 bp in length
* 48267 48366: gap of 100 bp
* 48367 49521: contig of 1155 bp in length
* 49522 49621: gap of 100 bp
* 49622 50802: contig of 1181 bp in length
* 50803 52050: contig of 1148 bp in length
* 52051 52150: gap of 100 bp
* 52151 53133: contig of 1163 bp in length
* 53134 53413: gap of 100 bp
* 53414 54553: contig of 1140 bp in length
* 54554 54653: gap of 100 bp
* 54654 55769: contig of 1116 bp in length
* 55770 55869: gap of 100 bp
* 55870 57009: contig of 1140 bp in length
* 57010 57109: gap of 100 bp
* 57110 58261: contig of 1152 bp in length
* 58262 58361: gap of 100 bp
* 58362 59517: contig of 1156 bp in length
* 59518 59617: gap of 100 bp
* 59618 60805: contig of 1188 bp in length
* 60806 60905: gap of 100 bp
* 60906 62065: contig of 1160 bp in length
* 62066 62166: gap of 100 bp
* 62166 63347: contig of 1182 bp in length
* 63348 63447: gap of 100 bp
* 63448 64612: contig of 1165 bp in length
* 64613 64712: gap of 100 bp
* 64713 65883: contig of 1171 bp in length
* 65884 65983: gap of 100 bp
* 65984 67149: contig of 1166 bp in length
* 67150 67249: gap of 100 bp
* 67250 68394: contig of 1145 bp in length
* 68395 68494: gap of 100 bp
* 68495 69619: contig of 1125 bp in length
* 69620 69719: gap of 100 bp
* 69720 70858: contig of 1139 bp in length
* 70859 70958: gap of 100 bp
*
* Query Match 6.8%; Score 125; DB 14; Length 108992;
* Best Local Similarity 100.0%; Pred. No. 4.7e-59;
* Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* 425 AGGGTATGTCGGAGACCTTTGAGACGCTTGACACCTAGTCACAAAGGGGTCAAGGTGG 484
* |||||
```

```
Db 50305 AGGGTATGTCGGAGACCTTTGAGACGCTTGACACCTAGTCACAAAGGGGTCAAGGTGG 50364
Oy 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGA 544
|||
Db 50365 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGA 50424
Oy 545 AGCAG 549
|||
Db 50425 AGCAG 50429

RESULT 5
AC074220/c 89855 bp DNA linear HTG 15-MAY-2002
LOCUS Mus musculus strain C57BL6/J chromosome 17 clone RP23-76116,
DEFINITION WORKING DRAFT SEQUENCE, 38 unordered pieces.
AC074220
AC074220.7 GI:13876390
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 89855)
AUTHORS Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
High Throughput Mouse Sequencing
Unpublished
REFERENCE 2 (bases 1 to 89855)
AUTHORS Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
TITLE Submitted (20-JUL-2000) Department of Molecular Genetics, Albert
JOURNAL Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
AUTHORS Bronx, NY 10461, USA
COMMENT On Apr 28, 2001 this sequence version replaced gi:11276104.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcg.org/Sequence/mouse.html
Contact: hpcg@mendel.mgh.harvard.edu
-----Summary Statistics
Center project name: ADY
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 79355 at least Q20
*Consensus quality: 75365 at least Q30
*Consensus quality: 69022 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 89115 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 7962: contig of 7962 bp in length
* 7963 7982: gap of unknown length
* 7983 13924: contig of 5942 bp in length
* 13925 13944: gap of unknown length
* 13945 18990: contig of 5046 bp in length
* 18991 18010: gap of unknown length
* 19011 22501: contig of 3491 bp in length
* 22502 22521: gap of unknown length
* 22522 27077: contig of 4556 bp in length
```

\* 27078 27097: gap of unknown length  
\* 27098 28819: contig of 1722 bp in length  
\* 28820 28839: gap of unknown length  
\* 28840 31224: contig of 2385 bp in length  
\* 31225 31244: gap of unknown length  
\* 31245 35093: contig of 3849 bp in length  
\* 35094 35113: gap of unknown length  
\* 35114 38611: contig of 3498 bp in length  
\* 38612 38631: gap of unknown length  
\* 38632 42261: contig of 3630 bp in length  
\* 42262 42281: gap of unknown length  
\* 42282 44130: contig of 1849 bp in length  
\* 44131 44150: gap of unknown length  
\* 44151 47059: contig of 2909 bp in length  
\* 47060 47079: gap of unknown length  
\* 47080 49496: contig of 2417 bp in length  
\* 49497 49516: gap of unknown length  
\* 49517 52236: contig of 2720 bp in length  
\* 52237 52256: gap of unknown length  
\* 52257 54270: contig of 2014 bp in length  
\* 54271 54290: gap of unknown length  
\* 54291 56271: contig of 1981 bp in length  
\* 56272 58849: gap of unknown length  
\* 58850 58869: gap of unknown length  
\* 58870 60621: contig of 1752 bp in length  
\* 60622 60641: gap of unknown length  
\* 60642 62255: contig of 1614 bp in length  
\* 62256 62275: gap of unknown length  
\* 62276 63085: contig of 810 bp in length  
\* 63086 63105: gap of unknown length  
\* 63106 65440: contig of 2335 bp in length  
\* 65441 65460: gap of unknown length  
\* 65461 67629: contig of 2169 bp in length  
\* 67630 67649: gap of unknown length  
\* 67650 69178: contig of 1529 bp in length  
\* 69179 69198: gap of unknown length  
\* 69199 70095: contig of 897 bp in length  
\* 70096 70115: gap of unknown length  
\* 70116 71821: contig of 1706 bp in length  
\* 71822 71842: gap of unknown length  
\* 71843 73506: contig of 1665 bp in length  
\* 73507 73526: gap of unknown length  
\* 73527 74749: contig of 1223 bp in length  
\* 74750 74769: gap of unknown length  
\* 74770 75963: contig of 1194 bp in length  
\* 75964 75983: gap of unknown length  
\* 75984 77905: contig of 1922 bp in length  
\* 77906 77925: gap of unknown length  
\* 77926 79236: contig of 1311 bp in length  
\* 79237 79256: gap of unknown length  
\* 79257 81332: contig of 2076 bp in length  
\* 81333 81352: gap of unknown length  
\* 81353 83169: contig of 1817 bp in length  
\* 83170 83189: gap of unknown length  
\* 83190 84096: contig of 907 bp in length  
\* 84097 84116: gap of unknown length  
\* 84117 85749: contig of 1633 bp in length  
\* 85750 85769: gap of unknown length  
\* 85770 86770: contig of 1001 bp in length  
\* 86771 86790: gap of unknown length  
\* 86791 88207: contig of 1417 bp in length  
\* 88208 88227: gap of unknown length  
\* 88228 88976: contig of 749 bp in length  
\* 88977 88996: gap of unknown length  
\* 88997 89855: contig of 859 bp in length.

FEATURES  
source

1..89855  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="CS7BL6/J"  
/db\_xref="taxon:10090"  
/chromosome="17"

Query Match 5.4%; Score 99; DB 14; Length 89855;

Best Local Similarity 100.0%; Pred. No. 3.8e-44;

Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGAGGAAGCGCGCGGCTCTGCTCTGCTCGGCTCGGCGCATGGAGTCATGTC 68

/clone="RP23-76116"  
/sex="male"  
1..7962  
/note="assembly\_name:Contig145"  
7963..7982  
/estimated\_length=unknown  
7983..13924  
/note="assembly\_name:Contig144"  
13925..13944  
/estimated\_length=unknown  
13945..18990  
/note="assembly\_name:Contig143"  
18991..19010  
/estimated\_length=unknown  
19011..22501  
/note="assembly\_name:Contig142"  
22502..22521  
/estimated\_length=unknown  
22522..27077  
/note="assembly\_name:Contig141"  
27078..27097  
/estimated\_length=unknown  
27098..28819  
/note="assembly\_name:Contig140"  
28820..28839  
/estimated\_length=unknown  
28840..31224  
/note="assembly\_name:Contig139"  
31225..31244  
/estimated\_length=unknown  
31245..35093  
/note="assembly\_name:Contig138"  
35094..35113  
/estimated\_length=unknown  
35114..38611  
/note="assembly\_name:Contig137"  
38612..38631  
/estimated\_length=unknown  
38632..42261  
/note="assembly\_name:Contig136"  
42262..42281  
/estimated\_length=unknown  
42282..44130  
/note="assembly\_name:Contig135"  
44131..44150  
/estimated\_length=unknown  
44151..47059  
/note="assembly\_name:Contig134"  
47060..47079  
/estimated\_length=unknown  
47080..49496  
/note="assembly\_name:Contig133"  
49497..49516  
/estimated\_length=unknown  
49517..52236  
/note="assembly\_name:Contig132"  
52237..52256  
/estimated\_length=unknown  
52257..54270  
/note="assembly\_name:Contig131"  
54271..54290  
/estimated\_length=unknown  
54291..56271  
/note="assembly\_name:Contig130"  
56272..56291  
/estimated\_length=unknown  
56292..58849

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

gap

misc\_feature

6070 GGAGGAAGCGCGCGGCTCGCTCTGGGTCCGGCTGGGCGCATGGAGTCATGTC 6011  
 69 TGAGTCCGCGCCCGCTCCCTCTATTTCCTTTCGTGCT 107  
 6010 TGAGTCCGCGCCCGCTCCCTCTATTTCCTTTCGTGCT 5972

RESULT 6  
 AC134482 222256 bp DNA linear HTG 15-NOV-2002  
 LOCUS Rattus norvegicus clone CH230-179K4, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 2 unordered pieces.  
 AC134482  
 VERSION AC134482.2 GI:25007228  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Rattus.

1 (bases 1 to 222256)

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,  
 Allten,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregorgis,E., Gear,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,M., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensuwa,L., Loulsegod,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,  
 Popper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L.-L.,  
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasanu,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wlecyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.

Direct Submission  
 Unpublished

TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

2 (bases 1 to 222256)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (27-SEP-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

3 (bases 1 to 222256)  
 Rat Genome Sequencing Consortium.  
 Direct Submission

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 15, 2002 this sequence version replaced gi:23334632.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KCOV

Center clone name: CH230-179K4

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 190050 bases at least Q40

Consensus quality: 194119 bases at least Q30

Consensus quality: 196964 bases at least Q20

Estimated insert size: 204504; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved

\* 1 107899: contig of 107899 bp in length

\* 107900 107999: gap of unknown length

\* 108000 222256: contig of 114257 bp in length.

FEATURES

source

1..222256

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-179K4"

107900..107999

/estimated\_length=unknown

complement(113301..114059)

/note="clone boundary

clone\_end:Sp6

site:

end\_sequence:BH303478"

ORIGIN

Query Match 3.8%; Score 69; DB 14; Length 222256;

Best Local Similarity 100.0%; Pred. No. 6.1e-27;

Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 676 CGAGAGCGTGTCTGGCAGAGGGGACATAGCTCCTGGGAGGAGGAATCCCAAG 735  
 Db 130160 CGAGAGCGTGTCTGGCAGAGGGGACATAGCTCCTGGGAGGAGGAATCCCAAG 130219  
 QY 736 AAGAAGCGC 744  
 Db 130220 AAGAAGCGC 130228

RESULT 7  
 AC131887  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-3M14, WORKING DRAFT SEQUENCE, 3  
 AC131887  
 ACCESSION  
 VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE  
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Rattus;  
 1. (bases 1 to 246404)  
 Muzny D.Marie, Metzker M.Lee, Abranzen S., Adams C., Alder J.,  
 Allen C., Allen H., Albrooks S., Amin A., Anguiano D.,  
 Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,  
 Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,  
 Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,  
 Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,  
 Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,  
 Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,  
 Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,  
 Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,  
 Delgado O., Denson S., Derramo C., Ding Y., Dinh H., Divya K.,  
 Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,  
 Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,  
 Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,  
 Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,  
 Grebergeorgis E., Geer K., Gill R., Grady M., Guerra W., Guevara W.,  
 Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K.,  
 Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,  
 Hernandez R., Hines S., Hladun S.L., Hodgson A., Hogue M.,  
 Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,  
 Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,  
 Karpasch S., Kelly S., Khan Z., King L., Kovar C.,  
 Kwis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,  
 Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,  
 Lorensuhewa L., Loulseghe H., Lozado R.J., Lu X., Ma J.,  
 Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,  
 Mangum B., Mapua P., Martin K., Martin R., Martinez E.,  
 Mawhney S., McLeod M.P., McNeill T.Z., Meenen E.,  
 Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,  
 Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,  
 Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,  
 Nwakoelme O., Okwunogbo G., Olarnpunsagoon A., Pal S., Parks K.,  
 Pasternak S., Paul H., Perez A., Perez L., Pfankuch C.,  
 Plopper F., Poindexter A., Popovic D., Primus E., Pu L.-L.,  
 Puazo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,  
 Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,  
 Rivers C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,  
 Sanders W., Savery G., Scherer S., Scott G., Shatsman S., Shen H.,  
 Shetty J., Shvartsbeyn A., Sisson I., Sitter C.D., Smajda D.,  
 Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,  
 Steimle M., Strong R., Sutton A., Svatek A., Taber P., Taylor C.,  
 Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,  
 Valas R., Vera V., Villanana D., Waldron L., Walker B., Wang J.,  
 Wang Q., Wang S., Warren J., Warren R., Wei X., White F.,  
 Williams G., Willson R., Wleczyk R., Wooden H., Worley K.,  
 Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,  
 Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von  
 Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,  
 Weinstock G. and Gibbs R.A.  
 Direct Submission

TITLE

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
 2 (bases 1 to 246404)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (27-AUG-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 246404)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On May 9, 2003 this sequence version replaced gi:23270097.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBSQ

Center clone name: CH230-3M14

----- Summary Statistics

Assembly program: Atlas;

Consensus quality: 238268 bases at least Q40

Consensus quality: 240099 bases at least Q30

Consensus quality: 241516 bases at least Q20

Estimated insert size: 249377; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 161257: contig of 161257 bp in length

\* 161258 161357: gap of unknown length

\* 161358 188568: contig of 27211 bp in length

\* 188569 188668: gap of unknown length

\* 188669 246404: contig of 57736 bp in length.

Location/Qualifiers

FEATURES  
 source

1. 246404

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clones="CH230-3M14"

161258..161357

/estimated\_length=unknown

161358..163910

/note="wgs contig"

164279..166935

/note="wgs contig"

188569..188668

/estimated\_length=unknown

ORIGIN

```

Query Match      3.8%; Score 69; DB 14; Length 246404;
Best Local Similarity 100.0%; Pred. No. 6.2e-27;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 676 GCAGAGCGTGGTCTGGCAAGAGGGGACATACGCTCCCTGGAGGAGAAATCCAAAG 735
Db 23303 GCAGAGCGTGGTCTGGCAAGAGGGGACATACGCTCCCTGGAGGAGAAATCCAAAG 23362

Qy 736 AAGAAGCGC 744
Db 23363 AAGAAGCGC 23371

RESULT 8
AC135655/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-6908, *** SEQUENCING IN PROGRESS ***,
4 unordered pieces.
AC135655
AC135655.2 GI:25138854
HTG; HTGS_PHASE3; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
1 (bases 1 to 270958)
Munzy,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,I., Fan,G.,
Fernandez,S., Finley,M., Flag,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,I., Haviak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuwa,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
MaWhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackeleme,O., Okwou,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzio,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

```

```

Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 270958)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 270958)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:24158508.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be whole
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KCOMI
Center clone name: CH230-6908
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 206551 bases at least Q40
Consensus quality: 211776 bases at least Q30
Consensus quality: 215202 bases at least Q20
Estimated insert size: 208321; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 248554: contig of 248554 bp in length
* 248555 248654: gap of unknown length
* 248655 267829: contig of 19175 bp in length
* 267830 267929: gap of unknown length
* 267930 268946: contig of 1017 bp in length
* 268947 269046: gap of unknown length
* 269047 270958: contig of 1912 bp in length.
* Location/Qualifiers
* 1. 270958
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-6908"
* 1. 1141
* /note="wgs_contig"
* 91405. 94036
* /note="wgs_contig"
FEATURES
source
misc_feature
misc_feature

```



misc_feature	473	/note="n includes a, c, g, or t, or contains no nucleotide"	
misc_feature	475	/note="n includes a, c, g, or t, or contains no nucleotide"	
misc_feature	477	/note="n includes a, c, g, or t, or contains no nucleotide"	
misc_feature	480	/note="n includes a, c, g, or t, or contains no nucleotide"	
misc_feature	482..483	/note="n includes a, c, g, or t, or contains no nucleotide"	
misc_feature	485..486	/note="n includes a, c, g, or t, or contains no nucleotide"	
misc_feature	488..490	/note="n includes a, c, g, or t, or contains no nucleotide"	
misc_feature	492..499	/note="n includes a, c, g, or t, or contains no nucleotide"	
misc_feature	501	/note="n includes a, c, g, or t, or contains no nucleotide"	
ORIGIN			
Query Match	2.7%;	Score 50; DB 6; Length 526;	
Best Local Similarity	100.0%;	Pred. No. 3.7e-16;	
Matches	50; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	586	GAGGACTGGTACAGGACCAACAGGAGGAGACCTGACTGAATTCCTCTG	635
Db	105	GAGGACTGGTACAGGACCAACAGGAGGAGACCTGACTGAATTCCTCTG	154
RESULT 10			
AX198865		580 bp DNA linear	PAT 29-AUG-2001
LOCUS	AX198865		
DEFINITION	Sequence 1320 from Patent WO0151513.		
ACCESSION	AX198865		
VERSION	AX198865.1	GI:15389191	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Algate, P.A.		
TITLE	Ovarian tumor-associated sequences		
JOURNAL	Patent: WO 015131-A 1320 19-JUL-2001;		
FEATURES	CORIXA CORPORATION (US)		
source	1..580	Location/Qualifiers	
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	2.7%;	Score 50; DB 6; Length 580;	
Best Local Similarity	100.0%;	Pred. No. 3.7e-16;	
Matches	50; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	586	GAGGACTGGTACAGGACCAACAGGAGGAGACCTGACTGAATTCCTCTG	635
Db	353	GAGGACTGGTACAGGACCAACAGGAGGAGACCTGACTGAATTCCTCTG	402
RESULT 11			
AX209392			



Hominidae: Homo  
1 (Baes 1977)

Strausberg, R.B., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L.H., Shennen, C.M., Schuler, G.D., Alschul, S.F., Zeeberg, B., Bueter, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Ditschenko, L., Marushina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellan, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fanej, J., Helton, E., Kettman, M., Madan, A.C., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bourdard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyewski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1277)

Director MGC Project.  
Direct Submission  
Submitted (22-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Guntaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAP Plate: 19 Row: n Column: 9.  
Location/Qualifiers  
1. 1277

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3457829"
/tissue_type="Cervix, carcinoma"
/clone_lib="Nih_MGC_12"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

```

```

      2.7%; Score 50; DB 8; Length 1277;
      Similarity 100.0%; Pred. No. 3.8e-16;
      50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
      6 GAGGACTGGTACAGGAACCCACCAGGAGGAAGACCTGACTGAAATTCCTCTG 635
      |||
      65 GAGGACTGGTACAGGAACCCACCAGGAGGAAGACCTGACTGAAATTCCTCTG 534
      |||
      eted: December 30, 2005, 07:12:39
      0085 secs

```

etted: December 30, 2005, 07:12:39  
0085 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 01:46:06 ; Search time 1051 Seconds  
(without alignments)  
11661.627 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacgaggaggagcgcc.....taaaaaaaaaaaaaaaaaaa 1839

Scoring table: OLIGO NUC  
Gapop 60.0, Gapext 60.0

Searched: 4996997 seqs, 332346308 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9953994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_21:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004as:\*
- 13: Geneseqn2004bs:\*
- 14: Geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1839	100.0	1839	4	AAS05278 Murine tr
2	1658	90.2	1848	4	AAS05286 DNA seque
3	1528	83.1	1553	14	Adv67441 Nucleotid
4	107	5.8	107	4	AAS05289 DNA seque
5	100	5.4	100	12	Adl88397 DNA up-re
6	100	5.4	100	12	Adl88396 DNA up-re
7	66	3.6	471	4	AAS05284 T243 gene
8	50	2.7	189	2	AAT23979 Human gen
9	50	2.7	526	13	ADU11888 Solid tum
10	50	2.7	580	4	AAS25051 Human ova
11	50	2.7	580	5	Ash83696 Human ova
12	50	2.7	901	8	ACC72109 Human NOV
13	50	2.7	934	12	ADM72164 Human NTR
14	50	2.7	1201	11	ADI30814 Human CDN
15	50	2.7	1201	13	ADS82881 Human lym
16	50	2.7	1226	13	ADUR2624 Human MDD
17	50	2.7	1362	4	AAS05279 Human tri
18	50	2.7	1378	10	ACA54677 Human NF-
19	50	2.7	1378	14	ADU83175 Human tri

20	50	2.7	1596	8	ACC72108	Acc72108 Human NOV
21	50	2.7	1675	4	AAS21311	Aas21311 Human CDN
22	50	2.7	1675	8	ACA03670	Aca03670 cDNA enco
23	50	2.7	1675	8	ABX89208	Abx89208 DNA enco
24	50	2.7	1675	8	ACD41862	Ac41862 Human sec
25	50	2.7	1675	8	ACA04091	Aca04091 Human CDN
26	50	2.7	1675	9	ADA45654	Ada45654 Novel hum
27	50	2.7	1675	9	ADA76085	Ada76085 Human PRO
28	50	2.7	1675	9	ADA18735	Ada18735 Human PRO
29	50	2.7	1675	9	ADA61358	Ada61358 Homo sapi
30	50	2.7	1675	9	ADB19143	Adb19143 Novel hum
31	50	2.7	1675	9	ADB27684	Adb27684 cDNA enco
32	50	2.7	1675	9	ADA86163	Ada86163 Novel hum
33	50	2.7	1675	9	ADB15727	Adb15727 Human PRO
34	50	2.7	1675	9	ADA47513	Ada47513 Human PRO
35	50	2.7	1675	9	ADA67308	Ada67308 Human PRO
36	50	2.7	1675	9	ADB30315	Adb30315 cDNA enco
37	50	2.7	1675	9	ADA85611	Ada85611 Novel hum
38	50	2.7	1675	9	ADA96823	Ada96823 Human PRO
39	50	2.7	1675	9	ADA79127	Ada79127 Human PRO
40	50	2.7	1675	9	ADA87266	Ada87266 Novel hum
41	50	2.7	1675	9	ADB16468	Adb16468 Human PRO
42	50	2.7	1675	9	ADA91560	Ada91560 Novel hum
43	50	2.7	1675	9	ADB14623	Adb14623 Human PRO
44	50	2.7	1675	9	ADB18584	Adb18584 Novel hum
45	50	2.7	1675	9	ADA93799	Ada93799 Human PRO

## ALIGNMENTS

### RESULT 1

AAS05278  
ID AAS05278 standard; cDNA; 1839 BP.

XX AAS05278;

DT 07-SEP-2001 (first entry)

DE Murine trinucleotide repeat protein (TRP) cDNA sequence.

KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;  
KW transgenic animal; knockout mouse; triplet repeat expansion;  
KW fragile X syndrome; Huntington's disease; mouse; murine; ss.

OS Mus sp.

FX Key

FT CDS

FT Location/Qualifiers

FT 55..885

FT /\*tag= a

FT /product= "TRP"

XX WO200130798-A1.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-US029382.

XX 26-OCT-1999; 99US-0161488P.

XX (DELT-) DELTAGEN INC.

XX Klein R, Matthews W, Moore M, Allen KD;

XX WPI; 2001-300473/31.

XX P-PSDB; AAU02498.

XX Novel transgenic animals useful as animal model for characterization of  
XX function of a gene encoding trinucleotide repeat proteins (TRPs),  
XX contains heterozygous disruption in a gene encoding TRP.

XX Claim 8; Fig 11; 106pp; English.











```

XX PN WO2003093445-A2.
XX PD 13-NOV-2003.
XX PF 05-MAY-2003; 2003WO-US0141114.
XX PR 03-MAY-2002; 2002US-0377383P.
XX PA (STOW-) STOWERS INST MEDICAL RES.
XX PI Li L;
XX DR WPI; 2004-022656/02.
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.
XX PS Claim 7; SEQ ID NO 4790; 123pp; English.
XX CC This invention relates to a novel method for predicting gene potential by
XX CC associating nucleic acid sequences of unknown function with particular
XX CC sub-population profiles. Specifically, it refers to classifying an
XX CC unknown multi-lineage affiliated gene by collecting hybridisation data to
XX CC develop a gene expression map, in order to determine the discrete sub-
XX CC population where it is expressed. The present invention describes methods
XX CC for predicting the lineage commitment of genes associated with the self-
XX CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX CC referred to as bone marrow stem cells populations. As such, these methods
XX CC can be used to identify associated multi-lineage affiliated genes and
XX CC hence the underlying molecular mechanisms in physiological haematopoietic
XX CC development. This polynucleotide sequence is DNA associated with a murine
XX CC HSC sub population of cells of the invention.
XX SQ Sequence 100 BP; 28 A; 23 C; 31 G; 18 T; 0 U; 0 Other;

Query Match 5.4%; Score 100; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.4e-25;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 513
DB 1 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 60

QY 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 553
DB 61 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 100

RESULT 6
ID ADL88396 standard; DNA; 100 BP.
XX AC ADL88396;
XX DT 20-MAY-2004 (first entry)
XX DE DNA up-regulated in murine haematopoietic stem cells SeqID 4789.
XX KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX KW HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
XX KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; ds.
XX OS Mus sp.
XX PN WO2003093445-A2.
XX PD 13-NOV-2003.
XX PF 05-MAY-2003; 2003WO-US0141114.
XX PR 03-MAY-2002; 2002US-0377383P.
XX PA (STOW-) STOWERS INST MEDICAL RES.
XX PI Li L;
XX DR WPI; 2004-022656/02.
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.
XX PS Claim 7; SEQ ID NO 4790; 123pp; English.
XX CC This invention relates to a novel method for predicting gene potential by
XX CC associating nucleic acid sequences of unknown function with particular
XX CC sub-population profiles. Specifically, it refers to classifying an
XX CC unknown multi-lineage affiliated gene by collecting hybridisation data to
XX CC develop a gene expression map, in order to determine the discrete sub-
XX CC population where it is expressed. The present invention describes methods
XX CC for predicting the lineage commitment of genes associated with the self-
XX CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX CC referred to as bone marrow stem cells populations. As such, these methods
XX CC can be used to identify associated multi-lineage affiliated genes and
XX CC hence the underlying molecular mechanisms in physiological haematopoietic
XX CC development. This polynucleotide sequence is DNA associated with a murine
XX CC HSC sub population of cells of the invention.
XX SQ Sequence 100 BP; 28 A; 23 C; 31 G; 18 T; 0 U; 0 Other;

Query Match 5.4%; Score 100; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.4e-25;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 513
DB 1 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 60

QY 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 553
DB 61 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 100

RESULT 7
ID AAS05284 standard; DNA; 471 BP.
XX AC AAS05284;
XX DT 07-SEP-2001 (first entry)
XX DE T243 gene homologous sequence #1 generated by PCR.
XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
XX KW transgenic animal; knockout mouse; triplet repeat expansion;
XX KW fragile X syndrome; Huntington's disease; mouse; ds.
XX OS Mus sp.
XX PN WO200130798-A1.
XX PD 03-MAY-2001.
XX PF 26-OCT-2000; 2000WO-US029382.
XX PR 26-OCT-1999; 99US-0161488P.
XX PA (DELT-) DELTAGEN INC.
XX PI Klein R, Matthews W, Moore M, Allen KD;
XX PR WPI; 2001-300473/31.
XX DR

```

```

PR 03-MAY-2002; 2002US-0377383P.
XX (STOW-) STOWERS INST MEDICAL RES.
XX PI Li L;
XX DR WPI; 2004-022656/02.
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.
XX PS Claim 7; SEQ ID NO 4789; 123pp; English.
XX CC This invention relates to a novel method for predicting gene potential by
XX CC associating nucleic acid sequences of unknown function with particular
XX CC sub-population profiles. Specifically, it refers to classifying an
XX CC unknown multi-lineage affiliated gene by collecting hybridisation data to
XX CC develop a gene expression map, in order to determine the discrete sub-
XX CC population where it is expressed. The present invention describes methods
XX CC for predicting the lineage commitment of genes associated with the self-
XX CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX CC referred to as bone marrow stem cells populations. As such, these methods
XX CC can be used to identify associated multi-lineage affiliated genes and
XX CC hence the underlying molecular mechanisms in physiological haematopoietic
XX CC development. This polynucleotide sequence is DNA associated with a murine
XX CC HSC sub population of cells of the invention.
XX SQ Sequence 100 BP; 28 A; 23 C; 31 G; 18 T; 0 U; 0 Other;

Query Match 5.4%; Score 100; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.4e-25;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 513
DB 1 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 60

QY 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 553
DB 61 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 100

RESULT 7
ID AAS05284 standard; DNA; 471 BP.
XX AC AAS05284;
XX DT 07-SEP-2001 (first entry)
XX DE T243 gene homologous sequence #1 generated by PCR.
XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
XX KW transgenic animal; knockout mouse; triplet repeat expansion;
XX KW fragile X syndrome; Huntington's disease; mouse; ds.
XX OS Mus sp.
XX PN WO200130798-A1.
XX PD 03-MAY-2001.
XX PF 26-OCT-2000; 2000WO-US029382.
XX PR 26-OCT-1999; 99US-0161488P.
XX PA (DELT-) DELTAGEN INC.
XX PI Klein R, Matthews W, Moore M, Allen KD;
XX PR WPI; 2001-300473/31.
XX DR

```







DR P-PSDB; ADM72134.

XX New human neurotransmission-associated proteins and polynucleotides for

PT diagnosing, preventing or treating diseases or conditions associated with

PT aberrant protein expression, e.g. cancer, hepatitis, AIDS, obesity or

PT stroke.

XX Claim 5; SEQ ID NO 45; 216pp; English.

XX The invention relates to human neurotransmission-associated proteins

CC (NTRAN) and encoding polynucleotides. The NTRAN polypeptides can be

CC expressed by standard recombinant methodology. The polypeptides and

CC polynucleotides are useful in diagnosing, preventing or treating diseases

CC or conditions associated with the decreased expression or overexpression

CC of NTRAN, such as autoimmune/inflammatory, cardiovascular, neurological,

CC developmental, cell proliferative, transport, psychiatric, metabolic or

CC endocrine disorders. These diseases may include AIDS, allergies, atopic

CC dermatitis, arthritis, thyroiditis, obesity, Parkinson's disease,

CC Alzheimer's disease, stroke, epilepsy, myocardial infarction,

CC hypertension, cancer, atherosclerosis or hepatitis. These are also useful

CC in assessing the effects of exogenous compounds on the expression of

CC nucleic acid and amino acid sequences of NTRAN. The NTRAN or its

CC fragments are useful in screening compounds for effectiveness as agonist

CC or antagonist of the polypeptides, or in altering the expression of the

CC target polynucleotide and compounds that specifically bind to or modulate

CC the activity of the polypeptide. The microarray is useful in monitoring

CC or measuring protein-protein interactions, drug-target interactions, and

CC gene expression profiles. The present sequence represents a specific

CC example of a polynucleotide encoding a human NTRAN polypeptide.

XX Sequence 934 BP; 232 A; 248 C; 283 G; 171 T; 0 U; 0 Other;

XX

Query Match 2.7%; Score 50; DB 12; Length 934;

Best Local Similarity 100.0%; Pred. No. 4.7e-08;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 586 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAAATTCCTCTG 635

Db 595 GAGGACTGGTACAGGACACACAGGAGGAGACCTGACTGAAATTCCTCTG 644

|||||

RESULT 14

ADI30814

ID ADI30814 standard; cDNA; 1201 BP.

XX AC ADI30814;

XX

DE Human CDNA #140.

XX Human; gene; ss; immunological response; immunopathological condition;

KW Crohn's disease; asthma; ulcerative colitis; hyperesoinophilia;

KW irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;

KW acute monocytic leukaemia; antiinflammatory; antiaesthetic; antiulcer;

KW osteopathic; antiarthritic; antirheumatic; cytostatic.

XX

OS Homo sapiens.

XX US6607879-B1.

PN 19-AUG-2003.

XX

XX 09-FEB-1998; 98US-00023655.

PF

XX 09-FEB-1998; 98US-00023655.

PR

XX (INCY-) INCYTE CORP.

PA

XX Cocks BG, Stuart SG, Seilhamer JJ;

PI

XX WPI; 2003-895307/82.

DR

XX

PT A composition comprising a plurality of cDNAs, useful for detecting

PT altered expression of genes in an immunological response or for

PT diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma

PT or osteoarthritis.

XX Claim 1; SEQ ID NO 140; 50pp; English.

XX The invention relates to a composition comprising a plurality of cDNAs

CC for detecting the altered expression of genes in an immunological

CC response. The invention also relates to a method of diagnosing or

CC monitoring the treatment of an immunopathological condition in a sample,

CC comprising obtaining nucleic acids from a sample, contacting the nucleic

CC acids of the sample with an array comprising the plurality of cDNAs under

CC conditions to form one or more hybridisation complexes, detecting the

CC hybridisation complexes and comparing the levels of the detected

CC hybridisation complexes with the level of hybridisation complexes

CC detected in a non-diseased sample, where an altered level of the detected

CC hybridisation complexes correlates with the presence of an

CC immunopathological condition. Also disclosed are an expression profile

CC comprising a microarray and a plurality of detectable complexes and a

CC method for identifying a plurality of polynucleotide probes. The cDNAs

CC are useful as hybridisable array elements in a microarray for monitoring

CC the expression of target polynucleotides. The microarray can be used in

CC the diagnosis of an immunopathology, such as Crohn's disease, asthma,

CC ulcerative colitis, hyperesoinophilia, irritable bowel syndrome,

CC osteoarthritis, rheumatoid arthritis or acute monocytic leukaemia, and in

CC identifying agents for the treatment of the diseases. The microarray may

CC also be used in drug discovery and development, toxicological and

CC carcinogenicity studies, forensics or pharmacogenomics. The composition

CC may also be used in purification of a subpopulation of mRNAs, cDNAs or

CC genomic fragments. This sequence represents a human cDNA of the

CC invention. Note: The sequence data for this patent did not form part of

CC the printed specification but was obtained in electronic format directly

CC from USPTO at seqdata.uspto.gov/sequence.html.

XX

Sequence 1201 BP; 293 A; 342 C; 347 G; 219 T; 0 U; 0 Other;

XX

Query Match 2.7%; Score 50; DB 11; Length 1201;

Best Local Similarity 100.0%; Pred. No. 4.5e-08;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 586 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAAATTCCTCTG 635

Db 421 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAAATTCCTCTG 470

|||||

RESULT 15

ADS82881

ID ADS82881 standard; cDNA; 1201 BP.

XX AC ADS82881;

XX

DT 11-AUG-2005 (first entry)

XX Human lymph node cDNA #140.

DE

XX ss; gene; human; immunological response; blood cell; cancer;

KW immunopathological; AIDS; allergy; anaemia; asthma; atherosclerosis;

KW bronchitis; ulcerative colitis; diabetes; multiple sclerosis;

KW osteoporosis; pancreatitis; infection; arthritis; lymph node.

XX

OS Homo sapiens.

XX US2004077003-A1.

PN 22-APR-2004.

XX

XX 14-AUG-2003; 2003US-00641643.

PF

XX 09-FEB-1998; 98US-00023655.

PR

XX (INCY-) INCYTE CORP.

PA

Search completed: December 30, 2005, 04:41:05  
Job time : 1054 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 03:44:32 ; Search time 6729 Seconds  
(without alignments)  
12786.661 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacgaggaggagcgc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 41078325 seqs, 23393541228 residues

Word size : 0

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:\*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_hc:*
5:	gb_est4:*
6:	gb_est5:*
7:	gb_est6:*
8:	gb_est7:*
9:	gb_ges1:*
10:	gb_ges2:*
11:	gb_ges3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1785	97.1	1787	4 AK086940	AK086940 Mus muscu
2	1690	91.9	1832	4 AK049358	AK049358 Mus muscu
3	1683	91.5	1684	4 AK031742	AK031742 Mus muscu
4	1609	87.5	1896	4 AK085617	AK085617 Mus muscu
5	1528	83.1	1553	4 AK005532	AK005532 Mus muscu
6	1277	69.4	1621	4 AK0082340	AK0082340 Mus muscu
7	966	52.5	1376	4 AK082749	AK082749 Mus muscu
8	871	47.4	1586	4 AK010702	AK010702 Mus muscu
9	766	41.7	766	8 CX166520	CX166520 HES2 41
10	760	41.3	766	2 BG974240	BG974240 602843972
11	679	36.9	903	6 CB320984	CB320984 AGENCOURT
12	663	36.1	926	5 BQ922065	BQ922065 AGENCOURT
13	662	36.0	892	6 CF553710	CF553710 AGENCOURT
14	656	35.7	937	6 CB203728	CB203728 AGENCOURT
15	637	34.6	699	8 CX208584	CX208584 MNS15855
16	629	34.2	721	10 AY403772	AY403772 Mus muscu
17	626	34.0	627	2 BB531047	BB531047 BB531047
18	608	33.1	643	2 BB661603	BB661603 BB661603
19	601	32.7	625	2 BB633565	BB633565 BB633565
20	588	32.0	641	2 BB637089	BB637089 BB637089
21	586	31.9	586	3 BM232538	BM232538 K0324G03-
22	582	31.6	656	2 BI082275	BI082275 602877456

c

23	572	31.1	737	2	BG873094	BG873094
24	570	31.0	696	6	CD533185	CD533185
25	565	30.7	665	6	CF732471	CF732471
26	558	30.3	656	2	BB504341	BB504341
27	556	30.2	644	1	BB190512	BB190512
28	549	29.9	768	5	BUS22275	AGENCOURT
29	547	29.7	639	8	CX217093	CX217093
30	542	29.5	758	7	CO801920	AGENCOURT
31	541	29.4	623	2	BB661026	BB661026
32	539	29.3	836	2	BI078951	BI078951
33	539	29.3	1078	5	BQ714545	AGENCOURT
34	533	29.0	608	2	BB605812	BB605812
35	526	28.6	645	2	BB498921	BB498921
36	525	28.5	985	2	BF785328	BF785328
37	523	28.4	651	2	BB385488	BB385488
38	522	28.4	573	1	AA260027	va52c07.r
39	521	28.3	632	2	BB558801	BB558801
40	516	28.1	516	3	BM223931	K0143E06-
41	512	27.8	564	7	CN700300	E0433C02-
42	503	27.4	503	5	BU696150	LL2in1070
43	497	27.0	497	2	BG062828	L0958D12-
44	489	26.6	631	5	BY705743	BY705743
45	481	26.2	557	3	BM224022	K0144G10-

## ALIGNMENTS

RESULT 1	AK086940	1787 bp	linear	HTC 03-APR-2004
LOCUS	Mus musculus 0 day neonate lung cDNA, RIKEN full-length enriched library, clone:R030013D02 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.			
ACCESSION	AK086940			
VERSION	AK086940.1	GI:26352264		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	Carninci, P. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)			
PUBMED	10349636			
REFERENCE	1			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)			
PUBMED	11042159			
REFERENCE	2			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, K., Maemoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Izawa, Y., Tanaka, T., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer			
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)			
PUBMED	11076861			
REFERENCE	3			
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.			
TITLE	Functional annotation of a full-length mouse cDNA collection			
JOURNAL	Nature 409, 685-690 (2001)			
PUBMED	11076861			
REFERENCE	4			
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.			
TITLE	Functional annotation of a full-length mouse cDNA collection			
JOURNAL	Nature 409, 685-690 (2001)			
PUBMED	11076861			
REFERENCE	5			

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 1787)

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>

URL: <http://fantom.gsc.riken.jp/>.

FEATURES

source

1. .1787

Location/Qualifiers

organism="Mus musculus"

mol\_type="mRNA"

strain="C57BL/6J"

db\_xref="FANTOM,DB:E030013D02"

db\_xref="taxon:10090"

clone="E030013D02"

tissue type="lung"

clone\_libs="RIKEN full-length enriched mouse cDNA library"

dev\_stage="0 day neonate"

22. .852

/note="unnamed protein product; 1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE) (SPTR|Q9DAU1, evidence: FASTY, 100%ID, 100%length, match=828)

putative"

/codon\_start=1

/protein\_id="BAC399769.1"

/db\_xref="GI:26352265"

/translation="MEGSELAPCLLPILLPLLLPAPKLPSPGPAEETDWR LPSCEVKYAVELKSAFETGKTKEVIDTGYLIDKGSGVKYKSDLRLIEVET ICKRLLYSLHKERTGSRNPAKMSFETFLHNLVHKGVKVMVDIPYELNETSABVA DLKQCDVLVEEFBEVIEDLVRNHEEDLTFELCANHLVKGKUTSCLAERWSGKGD I ASLGGKSKKRSKGVKSGSSGSKQRKELOGLEDANAEIEEGVQKASPLPHSPDEL "

CDS

Query Match 97.1%; Score 1785; DB 4; Length 1787;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1785; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 GCTCTGGGTCGGGTCGGCCATGGAGTCCATGTCTGAGCTCGCGCCCGGCTGCTTTAT 94

Db 2 GCTCTGGGTCGGGTCGGCCATGGAGTCCATGTCTGAGCTCGCGCCCGGCTGCTTTAT 61

Qy 95 TTCTTTTGTCTGCTGCTTCCGCTGCTGCTCCTTCTGCTCCGCCGGAAGCTAGGCCGAGTC 154

Db 62 TTCTTTTGTCTGCTGCTTCCGCTGCTGCTCCTTCTGCTCCGCCGGAAGCTAGGCCGAGTC 121

ORIGIN

1155 CCGCGGGGCTGAGGAGACCGGACTGGGTGCGATTGCCAGCAAAATGCCAAAGTGTGCAAGT 214

1122 CCGCGGGGCTGAGGAGACCGGACTGGGTGCGATTGCCAGCAAAATGCCAAAGTGTGCAAGT 181

215 ATGTTGCTGTGGAGCTGAAGTCGGCTTTTGTAGGAAACGGGAAAGACCAAGGAAGTGTG 274

182 ATGTTGCTGTGGAGCTGAAGTCGGCTTTTGTAGGAAACGGGAAAGACCAAGGAAGTGTG 241

275 ACACGGCTATGGCATCTCGACCGGAGGGCTCTGGAGTCAAGTACACCAAGTGGACT 334

242 ACACGGCTATGGCATCTCGACCGGAGGGCTCTGGAGTCAAGTACACCAAGTGGACT 301

335 TACGGTTAAATTGAAGTCACTGAGACCAATTTTCAAGAGGGCTTCTGGACTACAGCTGCACA 394

302 TACGGTTAAATTGAAGTCACTGAGACCAATTTTCAAGAGGGCTTCTGGACTACAGCTGCACA 361

395 AGGAGAGACTGGCAGCAACCGGTTTTCGCAAGGGTATGTTCGGAGACCTTTGAGAGCTGTC 454

362 AGGAGAGACTGGCAGCAACCGGTTTTCGCAAGGGTATGTTCGGAGACCTTTGAGAGCTGTC 421

455 ACAAACCTAGTCCCAAAAGGGTCAAGTGTGTGATGGATATCCCTATGAGCTGTGGAACG 514

422 ACAAACCTAGTCCCAAAAGGGTCAAGTGTGTGATGGATATCCCTATGAGCTGTGGAACG 481

515 AGACCTCAGCAGAGGTGGCTGACCTCAAGAACAGTGTGACGTGCTGGTGGAAAGTGTG 574

482 AGACCTCAGCAGAGGTGGCTGACCTCAAGAACAGTGTGACGTGCTGGTGGAAAGTGTG 541

575 AAGAGGTGATTGAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAAATTCCTCT 634

542 AAGAGGTGATTGAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAAATTCCTCT 601

635 GTGCCAACCACTGCTGAGAGGAAGACACGAGTTCGCTAGCAGAGCGGTGCTTGCGCA 694

602 GTGCCAACCACTGCTGAGAGGAAGACACGAGTTCGCTAGCAGAGCGGTGCTTGCGCA 661

695 AGAAGGGGGACATAGCTCTCCCTGGGAGGAAGAAATCCAAAGAAAGCGCAGCGAGTCA 754

662 AGAAGGGGGACATAGCTCTCCCTGGGAGGAAGAAATCCAAAGAAAGCGCAGCGAGTCA 721

755 AGGCTCTCTCAGTGGCAGCAGCAAGCAGAGGAAGAACTGGGGGGCTGGGGGAGGATG 814

722 AGGCTCTCTCAGTGGCAGCAGCAAGCAGAGGAAGAACTGGGGGGCTGGGGGAGGATG 781

815 CCACGCGGAGGAGGAGGGTGTGACAGAGGATCGCCCTCCACACAGCCCCCTG 874

782 CCACGCGGAGGAGGAGGGTGTGACAGAGGATCGCCCTCCACACAGCCCCCTG 841

875 ATGAGCTGTGAGCCAGCTTAGTGTCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934

842 ATGAGCTGTGAGCCAGCTTAGTGTCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 901

935 CACGACACGCGCAGCGCAGCGCAGCTCCAGCAAGGACAGCTGTGTTCAGCATCAGGTCT 994

902 CACGACACGCGCAGCGCAGCGCAGCTCCAGCAAGGACAGCTGTGTTCAGCATCAGGTCT 961

995 CTTCCCTTGGTGTGCCCCCTTTTCCCTTCCCTTGAACACAGCAAGGTTGGAGGATCTGG 1054

962 CTTCCCTTGGTGTGCCCCCTTTTCCCTTCCCTTGAACACAGCAAGGTTGGAGGATCTGG 1021

1055 GGTCTGGGAGACGGCACCCCAAGGAAGAGGAGGAGGACAGCAAGGACGCTCTCTTTC 1114

1022 GGTCTGGGAGACGGCACCCCAAGGAAGAGGAGGAGGAGGACAGCAAGGACGCTCTCTTTC 1081

1115 TACACAGTCCCTTACGAGCTCCGGGTCCACCCAGCATCCCGAGGCTGAGATCCAGGC 1174

1082 TACACAGTCCCTTACGAGCTCCGGGTCCACCCAGCATCCCGAGGCTGAGATCCAGGC 1141

1175 TCCTGACATGGAAGCTGAGAGCATGAGGCACATAAGATGCTCACCGGCCCTTTCAG 1234

1142 TCCTGACATGGAAGCTGGAAGCATGAGGCACATAAGATGCTCACCGGCCCTTTCAG 1201





2  
REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sasaki, N., Carninci, P., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
PUBMED 11042159  
REFERENCE  
AUTHORS 3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
PUBMED 11076861  
REFERENCE  
AUTHORS 4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
REFERENCE  
AUTHORS 5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
REFERENCE  
AUTHORS 6 (bases 1 to 1684)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukaya, S., Furuno, M., Hanaoka, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
FEATURES  
source  
1. 1684  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:6230400B22"  
/db\_xref="taxon:10090"  
/clone="6230400B22"  
/issue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="11 days embryo"  
1. 757  
/note="unnamed protein product; 1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA

1600025D17 GENE) (SPTR|Q9DAU1, evidence: FASTY, 100%ID, 100%length, match=828)  
putative"  
/protein id="BAC27534.1"  
/db\_xref="GI:26327581"  
/translation="PAPKGPSAGABETDMVRLPSKCEVKYVAVELKSAPEETGKT  
KEVIDTGVGLDGKSGVKYKSDRLIEVETETICKRLDLSLHKERTGNSRPAKMS  
ETTFELHNLVHKGVKVMNDIPYELWNESAEDLVKQCDLVEFEVIEDWYRNHQ  
BEDTETFLCANHLVKGDTSCLAERWSGKGDIASLGGKKKXKXKXSGVKGSSGSSKQ  
RKELGGLGEDANABEEGVQKASPLPHSPDEL"  
ORIGIN  
Query Match 91.5%; Score 1683; DB 4; Length 1684;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 130 CCTGCCCGAAGTAGGCCCGAGTCCCGCGGGCTGAGGAGACCGACTCGGTGCGATTG 189  
DB 2 CCTGCCCGAAGTAGGCCCGAGTCCCGCGGGCTGAGGAGACCGACTCGGTGCGATTG 61  
QY 190 CCCAGCAATGCGAAGTGTGCAAGTATGTTGTTGGAGCTGAAGTCCGCTTTTTCAGGAA 249  
DB 62 CCCAGCAATGCGAAGTGTGCAAGTATGTTGTTGGAGCTGAAGTCCGCTTTTTCAGGAA 121  
QY 250 ACGGAAAGACCAAGGAAGTATTGACACCGGCTATGGCATCTCTGGACGGGAAGGGCTCT 309  
DB 122 ACGGAAAGACCAAGGAAGTATTGACACCGGCTATGGCATCTCTGGACGGGAAGGGCTCT 181  
QY 310 GGAGTCAAGTACACCAAGTCCGACTTACGGTTTATTGAAGTCACTGAGACCAATTTGCAAG 369  
DB 182 GGAGTCAAGTACACCAAGTCCGACTTACGGTTTATTGAAGTCACTGAGACCAATTTGCAAG 241  
QY 370 AGGTTCTCGGACTACAGCTGTCACAAAGAGAGGAGTGGCAGCAACCGGTTTGCACAGGCT 429  
DB 242 AGGTTCTCGGACTACAGCTGTCACAAAGAGAGGAGTGGCAGCAACCGGTTTGCACAGGCT 301  
QY 430 ATGTGCGAGACCTTTGAGACGCTGCACAACTAGTTCACAAAGGGGTCAAGGTGGTGATG 489  
DB 302 ATGTGCGAGACCTTTGAGACGCTGCACAACTAGTTCACAAAGGGGTCAAGGTGGTGATG 361  
QY 490 GATATCCCTATGAGCTGTGAAACGAGACCTCAGCAGAGGTGGTGCATCTCAAGAACGAG 549  
DB 362 GATATCCCTATGAGCTGTGAAACGAGACCTCAGCAGAGGTGGTGCATCTCAAGAACGAG 421  
QY 550 TGTGACGTGCTGGTGAAGAGTGTGAAAGAGTGTGAGGACTGTTACAGAACACCCAG 609  
DB 422 TGTGACGTGCTGGTGAAGAGTGTGAAAGAGTGTGAGGACTGTTACAGAACACCCAG 481  
QY 610 GAGGAAGACCTTGACTGAATTCCTCTGTGCCCAACCACTGCTGAAGGGGAAAGACACGAGT 669  
DB 482 GAGGAAGACCTTGACTGAATTCCTCTGTGCCCAACCACTGCTGAAGGGGAAAGACACGAGT 541  
QY 670 TGCCTAGCAGAGCGGTGTCTGGCAAGAGGGGAGACATAGGCTTCCTGGAGGGAGGAAA 729  
DB 542 TGCCTAGCAGAGCGGTGTCTGGCAAGAGGGGAGACATAGGCTTCCTGGAGGGAGGAAA 601  
QY 730 TCCAAAGAGAGCGCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAACGAGGAGAG 789  
DB 602 TCCAAAGAGAGCGCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAACGAGGAGAG 661  
QY 790 GAACTGGGGGGCTTGGGGGAGGATGCCAAACCGCCGAGGAGGAGGGGTGTGCAGAGGCA 849  
DB 662 GAACTGGGGGGCTTGGGGGAGGATGCCAAACCGCCGAGGAGGAGGGGTGTGCAGAGGCA 721  
QY 850 TCGCCCTTCCACACAGCCCCCTTGATGAGTGTGAGCCAGCTTAGTGTCTTGAATCA 909  
DB 722 TCGCCCTTCCACACAGCCCCCTTGATGAGTGTGAGCCAGCTTAGTGTCTTGAATCA 781  
QY 910 AGACCCCTGACTTCAGAGCTTGGGACACGACGCGCAGCGCAGCGGAGCTCCAGCAGG 969  
DB 782 AGACCCCTGACTTCAGAGCTTGGGACACGACGCGCAGCGCAGCGGAGCTCCAGCAGG 841



```

/db_xref="taxon:10090"
/clone="D63004.7M13"
/tissue_type="kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
92_912
/note="unnamed protein product; 1600025D17RIK PROTEIN
(PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA
1600025D17 GENE) (SPTR|Q9DAUL, evidence: FASTY, 100%ID,
100%length, match=828)
putative"
/codon_start=1
/protein_id="BAC39489.1"
/db_xref="GI:26351705"
/translation="MESSELEAPRLCLPFLLLPPLPAPKLGSPAGABETDWR
LPSKEVCKYVALEKSAPEETGKTEVIDTGYGLDGGKGVKTKDLRLIETET
ICKRLDLSLHKEKTSNRPFAKMSSETFETLHNLVHKGVKVMVDI PYELMNETSAEVA
DLKKODVLVEFEFEEVIEDWRNHOREDLTFLCANHYLKGKDTSCLEARSGKGGDI
ASLGGKSKKRSRSGVKGSSSSSKORKELGGLGDANABEEGVQKASPLPHSPDDEL
"

```

CDS

ORIGIN

```

Query Match      87.5%; Score 1609; DB 4; Length 1896;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1729; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 9 GGAGGAAGCGCGCGGGTCCGCTCTGCTCTGGGTCCGGTCCGGCCATGGAGTCCATGTC 68
DB |
QY 36 GGAGGAAGCGCGCGGGTCCGCTCTGCTCTGGGTCCGGTCCGGCCATGGAGTCCATGTC 95
DB |
QY 69 TGAGCTCGGGCCCGCTGCTTATTTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 128
DB |
QY 96 TGAGCTCGGGCCCGCTGCTTATTTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 155
DB |
QY 129 TCCTGCCCGAGCTAGCGCGAGTCCCGCGGGCTGAGGAGCCGACTGGTCCGATT 188
DB |
QY 156 TCCTGCCCGAGCTAGCGCGAGTCCCGCGGGCTGAGGAGCCGACTGGTCCGATT 215
DB |
QY 189 GCCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAACTCGGCTTTTGAGGA 248
DB |
QY 216 GCCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAACTCGGCTTTTGAGGA 275
DB |
QY 249 AACCGGAAGACCAAGGAAGTATTGACACCGGCTATGCGATCTTGGACCGGAAGGGCTC 308
DB |
QY 276 AACCGGAAGACCAAGGAAGTATTGACACCGGCTATGCGATCTTGGACCGGAAGGGCTC 335
DB |
QY 309 TGGAGTCAAGTACCAAGTCCGACTAGGTTAATGACTGACTGAGACCAATTTGCAA 368
DB |
QY 336 TGGAGTCAAGTACCAAGTCCGACTAGGTTAATGACTGACTGAGACCAATTTGCAA 395
DB |
QY 369 GAGGCTTCTGGAAGTACAGCTGCAACAGGAGGAGTGGCAGCAACCGGTTTGCCAAAGG 428
DB |
QY 396 GAGGCTTCTGGAAGTACAGCTGCAACAGGAGGAGTGGCAGCAACCGGTTTGCCAAAGG 455
DB |
QY 429 TATGTCGAGACCTTTGAGACGCTGCAACACCTAGTCCAAAGGGGTCAGGTTGAT 488
DB |
QY 456 TATGTCGAGACCTTTGAGACGCTGCAACACCTAGTCCAAAGGGGTCAGGTTGAT 515
DB |
QY 489 GGATATCCCTTATGAGCTGTGGAACGAGACCTCAGCAGAGTGGTGGCTCAAGAAGCA 548
DB |
QY 516 GGATATCCCTTATGAGCTGTGGAACGAGACCTCAGCAGAGTGGTGGCTCAAGAAGCA 575
DB |
QY 549 GTGTGACGCTGTGTTGAAGAGTGTGTTGAAGAGTGTGTTGAAGAGTGTGTTGAAGAGT 608
DB |
QY 576 GTGTGACGCTGTGTTGAAGAGTGTGTTGAAGAGTGTGTTGAAGAGTGTGTTGAAGAGT 635
DB |
QY 609 GGAGGAAGACCTGACTGAAATTCCTCTGTGCCAACCAACCGTGTGAAAGGAAAGGACGAG 668
DB |
QY 636 GGAGGAAGACCTGACTGAAATTCCTCTGTGCCAACCAACCGTGTGAAAGGAAAGGACGAG 695
DB |
QY 669 TTGCTTACGAGCGGCTGTGCGCAAGAGGGGAGCATAGCTCCCTCGGGAGGGAAGAA 728
DB |
QY 696 TTGCTTACGAGCGGCTGTGCGCAAGAGGGGAGCATAGCTCCCTCGGGAGGGAAGAA 755

```

RESULT 5  
AK005532  
LOCUS

AK005532 1553 bp mRNA linear HTC 03-APR-2004

```

QY 729 ATCCAAAGAAAGCCAGCGAGTCAAGGGCTCTCTCAAGTGGCAGCAGCAAGCAGAGGAA 788
DB |
QY 756 ATCCAAAGAAAGCCAGCGAGTCAAGGGCTCTCTCAAGTGGCAGCAGCAAGCAGAGGAA 815
DB |
QY 789 GGAACTGGGGGCGCTGGGGGAGATGCCAAGCCGAGGAGGAGGGGTGTCAGAAAGGC 848
DB |
QY 816 GGAACTGGGGGCGCTGGGGGAGATGCCAAGCCGAGGAGGAGGGGTGTCAGAAAGGC 875
DB |
QY 849 ATCGCCCTCTCCACACAGCGCCCTCTGATGAGCTGTGAGCCACAGCTTAGTGTCTTGAATC 908
DB |
QY 876 ATCGCCCTCTCCACACAGCGCCCTCTGATGAGCTGTGAGCCACAGCTTAGTGTCTTGAATC 935
DB |
QY 909 AAGACCCCTGACTTCAGAGCTTTGGGACACGCAACAGCGGAGCGCGAGCGCTCCAGCAAG 968
DB |
QY 936 AAGACCCCTGACTTCAGAGCTTTGGGACACGCAACAGCGGAGCGCGAGCTCCAGCAAG 995
DB |
QY 969 GACAGCTCTGTCCAGACATCAGGCTCTCTCCCTTGGCTGTGCCCTTCTTCTTCCCTTGA 1028
DB |
QY 996 GACAGCTCTGTCCAGACATCAGGCTCTCTCCCTTGGCTGTGCCCTTCTTCTTCCCTTGA 1055
DB |
QY 1029 CAACAGCAAGAGGTGGAAGGATCTCGGGTGTCTGGGAGACGGCACCCCAAAGGAAAGGA 1088
DB |
QY 1056 CAACAGCAAGAGGTGGAAGGATCTCGGGTGTCTGGGAGACGGCACCCCAAAGGAAAGGA 1115
DB |
QY 1089 GGAGGAGCAGAAAGGAGCTCTCTTTTACAGAGTCCCTCTACAGAGTCCCGGGTCCACC 1148
DB |
QY 1116 GGAGGAGCAGAAAGGAGCTCTCTTTTACAGAGTCCCTCTACAGAGTCCCGGGTCCACC 1175
DB |
QY 1149 CAGCATCCCGAGGCTGAGATCCAGCTCTTGACATGGAAGCTGGAAGCATGAGGCACAT 1208
DB |
QY 1176 CAGCATCCCGAGGCTGAGATCCAGCTCTTGACATGGAAGCTGGAAGCATGAGGCACAT 1235
DB |
QY 1209 AAGATGCTCACAGGCGCCCTTACAGCAGGAAGGACTCCGTCGAGCTCAGCAGCCAGG 1268
DB |
QY 1236 AAGATGCTCACAGGCGCCCTTACAGCAGGAAGGACTCCGTCGAGCTCAGCAGCCAGG 1295
DB |
QY 1269 CTTGCTCTTCTTCCCAAGCAATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1328
DB |
QY 1296 CTTGCTCTTCTTCCCAAGCAATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
DB |
QY 1329 AGAACTTCAGGACAAACTCGGGTGTGGCAAAA-GGGGCTGGAGCGCCAGAGCCAGAGCC 1387
DB |
QY 1356 AGAACTTCAGGACAAACTCGGGTGTGGCAAAAAGGGGGCTGGAGCGCCAGAGCCAGAGCC 1415
DB |
QY 1388 ACAGCAGAGACTGAGAGAGGCGCACTGACCTAAACCCCTTGAAGGCAATCTGCAAGTT 1447
DB |
QY 1416 ACAGCAGAGACTGAGAGAGGCGCACTGACCTAAACCCCTTGAAGGCAATCTGCAAGTT 1475
DB |
QY 1448 CCGGTGTCACCCACTCTCTCTGAGGAGCGCTCATGCTGTGCCAGCCCTTCTCCAGGG 1507
DB |
QY 1476 CCGGTGTCACCCACTCTCTCTGAGGAGCGCTCATGCTGTGCCAGCCCTTCTCCAGGG 1535
DB |
QY 1508 CTACAGAGTAAACACCTTTTGGCTTTGGTTCCTGGGTCTCTCATCAGCTCCA 1567
DB |
QY 1536 CTACAGAGTAAACACCTTTTGGCTTTGGTTCCTGGGTCTCTCATCAGCTCCA 1595
DB |
QY 1568 GAGTGTCCCTCATCGATCTTTTGGCTTTTGGCTTGTCCCAATCCAGGGCTGGAAGGCCA 1627
DB |
QY 1596 GAGTGTCCCTCATCGATCTTTTGGCTTTTGGCTTGTCCCAATCCAGGGCTGGAAGGCCA 1655
DB |
QY 1628 TCACCATCATTTGGAGGCTTAACTGTCTAGTAACTAGAGAGTGTGGGAGCGCCCGGGTT 1687
DB |
QY 1656 TCACCATCATTTGGAGGCTTAACTGTCTAGTAACTAGAGAGTGTGGGAGCGCCCGGGTT 1715
DB |
QY 1688 GGTTTGGGTTAATCATCTACCTGCTCTCAGGCTTCTAAGTGTGGGAGCGCCCGGGTT 1737
DB |
QY 1716 GGTTTGGGTTAATCATCTACCTGCTCTCAGGCTTCTAAGTGTGGGAGCGCCCGGGTT 1765

```







QY 489 GGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGTGGCTGACCTCAAGAAGCA 548  
Db 506 GGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGTGGCTGACCTCAAGAAGCA 565  
QY 549 GTGTGACGCTGCTGGTGGAGAGCTTTGAAGAGGTGATTGAGGACTGTTACAGGAACCAACA 608  
Db 566 GTGTGACGCTGCTGGTGGAGAGCTTTGAAGAGGTGATTGAGGACTGTTACAGGAACCAACA 625  
QY 609 GGAGGAAGACCTGACTGTAATTCCTCTGTGCCAACCAACGCTGCTGAAGGGAAGGACACGAG 668  
Db 626 GGAGGAAGACCTGACTGTAATTCCTCTGTGCCAACCAACGCTGCTGAAGGGAAGGACACGAG 685  
QY 669 TTGCTTACGAGCGGTGCTGGCAAGAGGGGACATAGCTCTCCCTGGGAGGGAAGAA 728  
Db 686 TTGCTTACGAGCGGTGCTGGCAAGAGGGGACATAGCTCTCCCTGGGAGGGAAGAA 745  
QY 729 ATCCAAAGAAAGCCGAGCGAGTCAAGGGCTCTCCAGTGCAGCAGCAGCAGGAA 788  
Db 746 ATCCAAAGAAAGCCGAGCGAGTCAAGGGCTCTCCAGTGCAGCAGCAGCAGGAA 805  
QY 789 GGAATCTGGGGCTCTGGGGAGGATGTCACACCGCAGGAGGAGGGGTGTGCAAGGC 848  
Db 806 GGAATCTGGGGCTCTGGGGAGGATGTCACACCGCAGGAGGAGGGGTGTGCAAGGC 865  
QY 849 ATCGCCCTCCACACAGCCCTCTGATGAGCTGTGAGCCAGCTTAGTGTCTTGAATC 908  
Db 866 ATCGCCCTCCACACAGCCCTCTGATGAGCTGTGAGCCAGCTTAGTGTCTTGAATC 925  
QY 909 AAGACCCCTGACTTCAGAGCTTGGCAGCAGCAGCAGCGCAGCGCTCCAGCAAG 968  
Db 926 AAGACCCCTGACTTCAGAGCTTGGCAGCAGCAGCAGCGCAGCGCTCCAGCAAG 985  
QY 969 GACAGCTGTCTGACAGCTCAGGTCTCTCCCTTGGCTGTGCTTCTTCTTCTTCTTGA 1028  
Db 986 GACAGCTGTCTGACAGCTCAGGTCTCTCCCTTGGCTGTGCTTCTTCTTCTTCTTGA 1045  
QY 1029 CAACAGCAAGAGTGGAGAGTCTGGGTGTCTGGGAGCAGCGCACCCCAAGGGAAGAGA 1088  
Db 1046 CAACAGCAAGAGTGGAGAGTCTGGGTGTCTGGGAGCAGCGCACCCCAAGGGAAGAGA 1105  
QY 1089 GGAGAGCAGAGGAGCTCTCTTTACAGAGTCCCTCTCAGAGCTCCGGGTCCACC 1148  
Db 1106 GGAGAGCAGAGGAGCTCTCTTTACAGAGTCCCTCTCAGAGCTCCGGGTCCACC 1165  
QY 1149 CAGCATCCCGCTGAGATCAGGCTCTGACATGGAAGCTGAAGAGCATGAGGCAT 1208  
Db 1166 CAGCATCCCGCTGAGATCAGGCTCTGACATGGAAGCTGAAGAGCATGAGGCAT 1225  
QY 1209 AAGATGCTCACAGCGCCCTTTCAGCCAGGAAGGACTCCGTGCGAGCTTCAGCCAGG 1268  
Db 1226 AAGATGCTCACAGCGCCCTTTCAGCCAGGAAGGACTCCGTGCGAGCTTCAGCCAGG 1285  
QY 1269 CTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGCTCTTCTGCGATGGTAATTCG 1328  
Db 1286 CTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGCTCTTCTGCGATGGTAATTCG 1345  
QY 1329 AGAATCTCCAGGACAACTCGGGTGTGGCAAGAGGGCTGAGCCAGAGCCAGAGCCA 1388  
Db 1346 AGAATCTCCAGGACAACTCGGGTGTGGCAAGAGGGCTGAGCCAGAGCCAGAGCCA 1405  
QY 1389 GCCCAGAGATGCGAGAGGGGACCTGACCTAAACCCCTCTGGAAGCCAACTTCGAGTTC 1448  
Db 1406 GCCCAGAGATGCGAGAGGGGACCTGACCTAAACCCCTCTGGAAGCCAACTTCGAGTTC 1465  
QY 1449 CGGTGTCCACCCACTCTCTGAGAGCGCTCATGCTGTGCCAGCCCTTCTCCAGGGC 1508  
Db 1466 CGGTGTCCACCCACTCTCTGAGAGCGCTCATGCTGTGCCAGCCCTTCTCCAGGGC 1525  
QY 1509 TACCAGAGTAACACCTTTTGGCCCTTTC 1536  
Db 1526 TACCAGAGTAACACCTTTTGGCCCTTTC 1553

RESULT 6  
AK082340  
LOCUS  
DEFINITION  
Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone:C23003G09 product:1600025D17RIK PROTEIN  
(PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17  
GENE), full insert sequence.  
AK082340  
AK082340.1 GI:26349662  
HTC; CAP trapper.  
Mus musculus (house mouse)  
SOURCE  
ORGANISM  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
11076861  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 1621)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome

COMMENT







Best Local Similarity 99.4%; Pred. No. 0; Matches 1341; Conservative 0; Mismatches 7; Indels 1; Gaps 1;			
QY	443	TTGAGACGCTGACACACCTAGTCCAAAGAGGGGTCAAGGTGATGATATCCCTATG	502
Db	238	TTGAGACGCTGACACACCTAGTCCAAAGAGGGGTCAAGGTGATGATATCCCTATG	297
QY	503	AGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGACGTGCG	562
Db	298	AGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGACGTGCG	357
QY	563	TGGAAGAGTTTGAAGAGGTGATTGAGGACTGTTACAGGAACACACAGAGGAGACCTGA	622
Db	358	TGGAAGAGTTTGAAGAGGTGATTGAGGACTGTTACAGGAACACACAGAGGAGACCTGA	417
QY	623	CTGAATTTCTCTGTGTCACACCTGCTGAAAGGGGAAAGACACGAGTTGCCCTAGCAGAGC	682
Db	418	CTGAATTTCTCTGTGTCACACCTGCTGAAAGGGGAAAGACACGAGTTGCCCTAGCAGAGC	477
QY	683	GGTGTGTGGCAAGAGGGGACATAGCTTCCCTGGAGGGGAAAGAAATCCAAAGAAAGC	742
Db	478	GGTGTGTGGCAAGAGGGGACATAGCTTCCCTGGAGGGGAAAGAAATCCAAAGAAAGC	537
QY	743	GCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAGCAGAGGAGGAACTGGGGGGCC	802
Db	538	GCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAGCAGAGGAGGAACTGGGGGGCC	597
QY	803	TGGGGGAGGATGCCAAACGCGAGGAGGAGGGGTGTCAGAAAGCATCG-CCCCCTCCA	861
Db	598	TGGGGGAGGATGCCAAACGCGAGGAGGAGGGGTGTCAGAAAGCATCGCCCCCTCCA	657
QY	862	CACAGCCCCCTGATGAGTGTGAGCCAGCTTAGTGTCTTTGAATCAAGACCCCTGACT	921
Db	658	CACAGCCCCCTGATGAGTGTGAGCCAGCTTAGTGTCTTTGAATCAAGACCCCTGACT	717
QY	922	TCAGAGCTTGGACACGACACGCGCAGCGCAGCTCCAGCAAGGACAGCTGCTGTC	981
Db	718	TCAGAGCTTGGACACGACACGCGCAGCGCAGCTCCAGCAAGGACAGCTGCTGTC	777
QY	982	CAGCATCAGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1041
Db	778	CAGCATCAGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	837
QY	1042	TGGAAGGATCTGGGGTGTGGGAGACGCGACCCCAAGAGGAGAGGAGGAGGAGGAGGAG	1101
Db	838	TGGAAGGATCTGGGGTGTGGGAGACGCGACCCCAAGAGGAGAGGAGGAGGAGGAGGAG	897
QY	1102	GCAGCTCTCTTCTACAGTCCCTCCTCAGAGCTCCGGGTCCACCCAGCATCCCGAGG	1161
Db	898	GCAGCTCTCTTCTACAGTCCCTCCTCAGAGCTCCGGGTCCACCCAGCATCCCGAGG	957
QY	1162	CTGAGATCCAGGCTCTGACATGGAAGCTGAAGAGCATGAGGCACATAGATGCTCACA	1221
Db	958	CTGAGATCCAGGCTCTGACATGGAAGCTGAAGAGCATGAGGCACATAGATGCTCACA	1017
QY	1222	GGGCCCCCTTCAGCAGGAGGACTCCGTCAGAGCTCAGCAGCAGGCTGCTCTTCTCT	1281
Db	1018	GGGCCCCCTTCAGCAGGAGGACTCCGTCAGAGCTCAGCAGCAGGCTGCTCTTCTCT	1077
QY	1282	TCCACCAAGCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1341
Db	1078	TCCACCAAGCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1137
QY	1342	CAAACTCCGGTGTGACAAAGGGGCTGAGCCAGAGCCAGAGCCAGCCAGAGACTGC	1401
Db	1138	CAAACTCCGGTGTGACAAAGGGGCTGAGCCAGAGCCAGAGCCAGCCAGAGACTGC	1197
QY	1402	AGAGAGGGCCTGACCTAACCCCTCTGAAAGCCAAATCTGAGTTCCCGTGTCCACCCA	1461
Db	1198	AGAGAGGGCCTGACCTAACCCCTCTGAAAGCCAAATCTGAGTTCCCGTGTCCACCCA	1257
QY	1462	CTCTCTTGTGAGACGCTCATGCTCTGCCAGCCCTTCTCCAGGGCTACAGAGTAAC	1521

Db	1258	CTCCTCCTGAGGAGCGCCTCATGCTCTGCCAGCCCTTCTCCAGGGCTACAGAGTAAC	1317
QY	1522	ACCTTTTGGCCTTTGGTTTGGTTCTCTGGGTCTCTCATCAGAGTGTCCCTCAT	1581
Db	1318	ACCTTTTGGCCTTTGGTTTGGTTCTCTGGGTCTCTCATCAGAGTGTCCCTCAT	1377
QY	1582	CGATCTTTTGGCCTTTGGTTTGGTTCTCTGGGTCTCTCATCAGAGTGTCCCTCAT	1641
Db	1378	CGATCTTTTGGCCTTTGGTTTGGTTCTCTGGGTCTCTCATCAGAGTGTCCCTCAT	1437
QY	1642	GGCTTAACCTCTCAGTTACTAGGAGGTCTGGAGCGCCCGGGTGGTTGGGGTAATC	1701
Db	1438	GGCTTAACCTCTCAGTTACTAGGAGGTCTGGAGCGCCCGGGTGGTTGGGGTAATC	1497
QY	1702	ACTCATGGCTCTCAGCCTTTAAACATGACGAGCCCTTTAAATACAGTCTCTCTGTGG	1761
Db	1498	ACTCATGGCTCTCAGCCTTTAAACATGACGAGCCCTTTAAATACAGTCTCTCTGTGG	1557
QY	1762	TGACTCCACGCGCCCGCCACACACACCAT	1790
Db	1558	TGACTCCACGCGCCCGCCACACACACCAT	1586
RESULT 9			
CX166520			
LOCUS			
DEFINITION			
HES2C.41.D10.g1.A035 NIH MGC 258 Homo sapiens cdna clone			
IMAGE:7470381 5', mRNA sequence.			
ACCESSION			
CX166520			
VERSION			
CX166520.1 GI:56796600			
KEYWORDS			
EST.			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
Hominidae; Homo			
1 (Bases 1 to 766)			
REFERENCE			
NIH-MGC http://mgs.nci.nih.gov/.			
AUTHORS			
National Institutes of Health, Mammalian Gene Collection (MGC)			
TITLE			
Unpublished (1999)			
JOURNAL			
COMMENT			
Contact: Daniela S. Gerhard, Ph.D.			
Office of Cancer Genomics			
National Cancer Institute / NIH			
Bldg. 31 Rm10A07 Bethesda, MD 20892			
Email: cgaabs-remail.nih.gov			
Tissue Procurement: BresaGen, Inc.			
cDNA Library Preparation: Express Genomics, Inc.			
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			
Plate: LLAM15772 row: g column: 19			
Seq primer: JENREV (CAGGAACAGCTATGACC)			
High quality sequence stop: 766.			
FEATURES			
Location/Qualifiers			
1..766			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:7470381"			
/sex="male"			
/tissue_type="human embryonic stem cells differentiated to			
an early endodermal cell type"			
/cell_type="human embryonic stem cells"			
/cell_line="BG01"			
/lab_host="DH10B-T1 phage-resistant E. coli"			
/clone_lib="NIH MGC 258"			
/note="vector: pEXpress-1; Site 1: NotI; Site 2: EcoRV;			
RNA obtained from human embryonic stem cells isolated from			
the inner cell mass of blastocyst stage embryos and			
differentiated to an early endodermal cell type. Cell line			

id and NIH Registry designation is BG01. Positive for GAT34, Mix1, Mex1, HNP4alpha expression; negative for APP expression. Passage number 40. cDNA primed using oligo-dT primer: 5'-pGACTAGTCTAGATCGGAGCGGCCCTT25-3', and cloned into the EcorV/NotI sites of pExpress-1. This primary library is non-normalized (normalized primary library is NIH MGC 259). It was constructed by Express Genomics (Frederick, MD). Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA. Note: this is a Mammalian Gene Collection library."

## ORIGIN

Query Match 41.7%; Score 766; DB 8; Length 766;  
Best Local Similarity 100.0%; Pred. No. 5.7e-301; Mismatches 0; Indels 0; Gaps 0;  
Matches 766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 CTCTGGGTCCGGCTGGGCATGGAGTCCATGTCTGAGCTCGCGCCCGCTGCTCTTATT 95  
DB 1 CTCTGGGTCCGGCTGGGCATGGAGTCCATGTCTGAGCTCGCGCCCGCTGCTCTTATT 60

QY 96 TCTTTTGTGTGCTGCTTCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 155  
DB 61 TCTTTTGTGTGCTGCTTCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120

QY 156 CGCCGGGCTGAGGAGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 215  
DB 121 CGCCGGGCTGAGGAGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 180

QY 216 TGTCTCTGTGAGTCTGAAGTGGCTTTTGGAGAAACGGGAAAGACCAAGGAAGTGA 275  
DB 181 TGTCTCTGTGAGTCTGAAGTGGCTTTTGGAGAAACGGGAAAGACCAAGGAAGTGA 240

QY 276 CACCGGCTATGGCATCTCTGGACGGGAAGGGCTCTGGAGTCAAGTACACCAAGTCGACTT 335  
DB 241 CACCGGCTATGGCATCTCTGGACGGGAAGGGCTCTGGAGTCAAGTACACCAAGTCGACTT 300

QY 336 ACGGTTAATTGAAGTCACTGAGACCATTTGCAAGAGGCTCTGGAGTCAAGTCAAGTCAAA 395  
DB 301 ACGGTTAATTGAAGTCACTGAGACCATTTGCAAGAGGCTCTGGAGTCAAGTCAAGTCAAA 360

QY 396 GGAGAGGACTGGCAGCAACCGGTTTGCACAAAGGATGTCTGGAGACCTTTGAGACGCTGCA 455  
DB 361 GGAGAGGACTGGCAGCAACCGGTTTGCACAAAGGATGTCTGGAGACCTTTGAGACGCTGCA 420

QY 456 CAACCTAGTCCACAAAGGGGTCAAGTGGTGTATGATATFCCCTATGAGCTGTGGAAAGA 515  
DB 421 CAACCTAGTCCACAAAGGGGTCAAGTGGTGTATGATATFCCCTATGAGCTGTGGAAAGA 480

QY 516 GACCTCAGCAGAGTGGTGAACCTCAAGACGAGTGTGACGTGCTGGTGGAGAGTTTGA 575  
DB 481 GACCTCAGCAGAGTGGTGAACCTCAAGACGAGTGTGACGTGCTGGTGGAGAGTTTGA 540

QY 576 AGAGGTGATTGAGGACTGGTACAGAAACCAACAGGAGGACCTGACTGAATTCCTCTG 635  
DB 541 AGAGGTGATTGAGGACTGGTACAGAAACCAACAGGAGGAGACCTGACTGAATTCCTCTG 600

QY 636 TGCCCAACCACTGCTGAAGGGGAAAGGACACGAGTTGCCCTAGCAGAGCGGTGGTCTGGCAA 695  
DB 601 TGCCCAACCACTGCTGAAGGGGAAAGGACACGAGTTGCCCTAGCAGAGCGGTGGTCTGGCAA 660

QY 696 GAAGGGGACATAGCTCTCTGGAGGAGGAGAAATCCAGAGAGACCGCAGCGGAGTCAA 755  
DB 661 GAAGGGGACATAGCTCTCTGGAGGAGGAGAAATCCAGAGAGAGACCGCAGCGGAGTCAA 720

QY 756 GGGCTCTCTCAGTGGCAGCAGCAAGCAGAGGAGGAACTGGGGGGC 801  
DB 721 GGGCTCTCTCAGTGGCAGCAGCAGGAGGAGGAGAACTGGGGGGC 766

RESULT 10  
BG974240

LOCUS BG974240 766 bp mRNA linear EST 12-JUN-2001  
DEFINITION 602843972F1 NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:4979944 5',  
mRNA sequence.  
ACCESSION BG974240  
VERSION BG974240.1 GI:14361877  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 766)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth  
Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10978 row: 0 column: 17  
High quality sequence stop: 761.

FEATURES  
source

1..766  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4979944"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Mam4"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furth,  
NIH Reference for transgenic model: Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."

## ORIGIN

Query Match 41.3%; Score 760; DB 2; Length 766;  
Best Local Similarity 100.0%; Pred. No. 1.6e-298; Mismatches 0; Indels 0; Gaps 0;  
Matches 760; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 AAGGACACGAGTTGCCTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCCTCCCTG 717  
DB 1 AAGGACACGAGTTGCCTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCCTCCCTG 60

QY 718 GGAGGGAAGAAATCCAGAAGAAGCGCAGCGAGTCAAGGGGTCTCCAGTGGCAGCAGC 777  
DB 61 GGAGGGAAGAAATCCAGAAGAAGCGCAGCGAGTCAAGGGGTCTCCAGTGGCAGCAGC 120

QY 778 AAGCAGAGGAAGGAACCTGGGGGGCTGGGGAGGAGTGCACCGCCGAGAGGAGGGT 837  
DB 121 AAGCAGAGGAAGGAACCTGGGGGGCTGGGGAGGAGTGCACCGCCGAGAGGAGGGT 180

QY 838 GTGCAGAGGAGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGT 897  
DB 181 GTGCAGAGGAGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGT 240

QY 898 GTCTCTTGATCAAGACCCCTGACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCGCA 957  
DB 241 GTCTCTTGATCAAGACCCCTGACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCGCA 300

QY 958 GCTCAGCAGAGCAGCTGCTGTCCAGCATCAGGTCTCTCCCTGGCTGGCCCTTTC 1017  
DB 301 GCTCAGCAGAGCAGCTGCTGTCCAGCATCAGGTCTCTCCCTGGCTGGCCCTTTC 360





AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>;  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: L1AM13996 row: i column: 17  
 High quality sequence stop: 632.  
 Location/Qualifiers  
 1. .926  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:648664"  
 /tissue\_type="retina"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_94"  
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life technologies. Note: this is a NIH\_MGC Library."  
 ORIGIN  
 Query Match 36.1%; Score 663; DB 5; Length 926;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-259; Indels 0; Gaps 0;  
 Matches 663; Conservative 0; Mismatches 0;  
 Qy 541 AAGAGCAGTGTGACGTCTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGG 600  
 Db 1 AAGAGCAGTGTGACGTCTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGG 60  
 Qy 601 AACACACAGGAGGAAGACTGACTGAATTCCTGTGTGCAACACCGAGTCTCAAGGGAAG 660  
 Db 61 AACACACAGGAGGAAGACTGACTGAATTCCTGTGTGCAACACCGAGTCTCAAGGGAAG 120  
 Qy 661 GACACGAGTGTCTAGCAGAGCGGTGTCTGTGCAAGAGGGGACATAGCTCCTCTGGGA 720  
 Db 121 GACACGAGTGTCTAGCAGAGCGGTGTCTGTGCAAGAGGGGACATAGCTCCTCTGGGA 180  
 Qy 721 GGGAGAAATCCAAAGAAAGCGCAGCGAGGTCAAGGGCTCTCCAGTGGCAGCAGCAAG 780  
 Db 181 GGGAGAAATCCAAAGAAAGCGCAGCGAGGTCAAGGGCTCTCCAGTGGCAGCAGCAAG 240  
 Qy 781 CAGAGGAAGAACTGGGGGGCTGGGGAGAGTGCACCGCGAGGAGGAGGGGTGTG 840  
 Db 241 CAGAGGAAGAACTGGGGGGCTGGGGAGAGTGCACCGCGAGGAGGAGGGGTGTG 300  
 Qy 841 CAGAGGAGTACGCTCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGTCTC 900  
 Db 301 CAGAGGAGTACGCTCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGTCTC 360  
 Qy 901 CTTGAATCAAGACCCCTGACTTTGAGAGTCTGGGACACGACAGCGCAGCGCAGCT 960  
 Db 361 CTTGAATCAAGACCCCTGACTTTGAGAGTCTGGGACACGACAGCGCAGCGCAGCT 420  
 Qy 961 CCAGCAAGACAGCTGTGTTCAGAGTCTCCTCTCTGGTGTGTCCTTTTCCTT 1020  
 Db 421 CCAGCAAGACAGCTGTGTTCAGAGTCTCCTCTCTGGTGTGTCCTTTTCCTT 480  
 Qy 1021 CCCTTGAACACAGCAAGAGGTGGAAGATCTGGGGTCTGGGACACGCGACCCCAAGG 1080  
 Db 481 CCCTTGAACACAGCAAGAGGTGGAAGATCTGGGGTCTGGGACACGCGACCCCAAGG 540  
 Qy 1081 GAAGAGGAGGAGGAGCAGAGGAGCTCTCTTTTACACAGTCCCTTCACAGAGTCCGG 1140  
 Db 541 GAAGAGGAGGAGGAGCAGAGGAGCTCTCTTTTACACAGTCCCTTCACAGAGTCCGG 600

Qy 1141 GGTCCACCCAGCATCCCGAGCTGAGATCCAGGCTCCTGACATCGAAGCTGAAGAGCATG 1200  
 Db 601 GGTCCACCCAGCATCCCGAGCTGAGATCCAGGCTCCTGACATCGAAGCTGAAGAGCATG 660  
 Qy 1201 AGG 1203  
 Db 661 AGG 663  
 RESULT 13  
 CF553710  
 LOCUS  
 DEFINITION AGENCOURT 15614073 Organ of Corti Mus musculus cDNA clone  
 IMAGE:30519414 5', mRNA sequence.  
 ACCESSION CF553710  
 VERSION CF553710  
 KEYWORDS EST.  
 SOURCE CF553710.1 GI:34890544  
 ORGANISM Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 1 (bases 1 to 892)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Bechara Kachar and Celine Pompeia, NIDCD  
 cDNA Library Preparation: Bechara Kachar, Celine Pompeia, NIDCD  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: N1AM590 row: o column: 07  
 High quality sequence stop: 579.  
 Location/Qualifiers  
 1. .892  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30519414"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="Organ of Corti"  
 /note="Organ: Organ of Corti/Ear; Vector: pBluescriptSK-; Site 1: EcoRI; Site 2: XhoI; The organ of Corti (OC) was fine dissected from a total of 386 OC as follows: 102 samples from post-natal (P) day 5; 72 from P6; 60 from P7; 46 from P8; 18 from P9; 20 from P10; 14 from P12 and 24 from P13. Total RNA was extracted using the micro Fasttrack kit (catalog # K1593-02; Invitrogen, Carlsbad, CA), according to manufacturer's instructions. Reverse transcription and library construction were carried out with the Uni-Zap XR vector kit (catalog # 237211, Stratagene) and Uni-Zap XR Gigapack III Gold Cloning kit (catalog # 237612), both from Stratagene (La Jolla, CA, USA), according to manufacturer's instructions. The frequency distribution of the library is as follows: 72% of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and 0.1% 51-150. As to gene function, 45% of genes are present in GenBank and have known function; 23% have hits in GenBank, but do not have assigned function; 12% are uncharacterized ESTs and 20% are unidentified. Library created in the laboratory of M. Brownstein (NIMH, NIH). A complete library description can be found at ."  
 ORIGIN  
 Query Match 36.0%; Score 662; DB 6; Length 892;



Best Local Similarity 100.0%; Pred. No. 1.3e-258;		Matches 662; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	49	TGGGCCATGGAGTCCATGCTGAGCTCGCGCCCGCTGCTCTTATTTCCCTTTGCTGCTG	108
Db	9	TGGGCCATGGAGTCCATGCTGAGCTCGCGCCCGCTGCTCTTATTTCCCTTTGCTGCTG	68
Qy	109	CTGCTTCGCTGCTGCTCTCTCTGCTCCGCGGAGCTAGGCGCGAGTCCCGCGGGGCTGAG	168
Db	69	CTGCTTCGCTGCTGCTCTCTCTGCTCCGCGGAGCTAGGCGCGAGTCCCGCGGGGCTGAG	128
Qy	169	GAGACCGACTGGTGGGATGCGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGGAG	228
Db	129	GAGACCGACTGGTGGGATGCGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGGAG	188
Qy	229	CTGAAGTCGGCTTTTGGAGAAACGGGAAGACCAAGGAAGTGAATGACACCGGCTATGGC	288
Db	189	CTGAAGTCGGCTTTTGGAGAAACGGGAAGACCAAGGAAGTGAATGACACCGGCTATGGC	248
Qy	289	ATCTTGGACGGGAAGGGCTTGGAGTCAAGTACACAAAGTCGGAATTACGGTTAATTGAA	348
Db	249	ATCTTGGACGGGAAGGGCTTGGAGTCAAGTACACAAAGTCGGAATTACGGTTAATTGAA	308
Qy	349	GTCACTGAGACCAATTTGCAAGAGGCTTCTGGACTACGCTGCAACAGGAGGACTGCG	408
Db	309	GTCACTGAGACCAATTTGCAAGAGGCTTCTGGACTACGCTGCAACAGGAGGACTGCG	368
Qy	409	AGCAACCGGTTTCCCAAGGGTATGTCGAGAGCTTTTGAGAGCTGCTGCACAACTAGTCCAC	468
Db	369	AGCAACCGGTTTCCCAAGGGTATGTCGAGAGCTTTTGAGAGCTGCTGCACAACTAGTCCAC	428
Qy	469	AAAGGGGTCAAGTGGTGAATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAG	528
Db	429	AAAGGGGTCAAGTGGTGAATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAG	488
Qy	529	GTGGTGACCTCNAGAACAGTGTACCTGCTGCTGGAGAGTGTGAAGAGTGAATTGAG	588
Db	489	GTGGTGACCTCNAGAACAGTGTACCTGCTGCTGGAGAGTGTGAAGAGTGAATTGAG	548
Qy	589	GACTGGTACAGGAACCAACAGGAGGAACCTGACTGAATTCCTCTGTGCCAACACCACTG	648
Db	549	GACTGGTACAGGAACCAACAGGAGGAACCTGACTGAATTCCTCTGTGCCAACACCACTG	608
Qy	649	CTGAAGGAAAGGACACAGGTTGCTCTAGCAGCGGTGTCTGGCAAGAGGGGACATA	708
Db	609	CTGAAGGAAAGGACACAGGTTGCTCTAGCAGCGGTGTCTGGCAAGAGGGGACATA	668
Qy	709	GC 710	
Db	669	GC 670	

RESULT 14  
CB203728  
LOCUS  
DEFINITION  
IMAGE:30142422 5', mRNA sequence.  
ACCESSION  
CB203728  
VERSION  
CB203728.1  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 937)  
NIH-MGC http://mgc.nci.nih.gov/.  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDA00051 row: k column: 07  
High quality sequence stop: 609.  
Location/Qualifiers  
1. 937  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30142422"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_135"  
Note="Vector: pCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized full-length enriched library from pooled mouse  
embryonic limb, maxilla and mandible, day 12.5, 13.5,  
14.5, and 15.5 (size selected for the 0.5-1 kb fragments)  
Cloned directionally, priming method: Oligo-dT. cDNA  
enrichment: >1k bp. Average insert size 1.6k bp.  
Normalization (Cot value): 7.5 kb. Priming sequence:  
5'GACTAGTCTAGATCGAGCGCGCCCT(T)3' Tissue contributed  
by, David Rowe. Library constructed by ResGen, Invitrogen  
Corp."

ORIGIN

Query Match 35.7%; Score 656; DB 6; Length 937;		Best Local Similarity 99.9%; Pred. No. 3.6e-256;		Matches 706; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	559	CTGTGGGAAGAGTTTGAAGAGGTGATTGAGACTGTTACAGGAACCAACAGGAGGAAGAC	618		
Db	1	CTGTGGGAAGAGTTTGAAGAGGTGATTGAGACTGTTACAGGAACCAACAGGAGGAAGAC	60		
Qy	619	CTGACTGAATTCCTCTGTGCCAACACCGTGTGAGGGAAGGACACGAGTTGCCTAGCA	678		
Db	61	CTGACTGAATTCCTCTGTGCCAACACCGTGTGAGGGAAGGACACGAGTTGCCTAGCA	120		
Qy	679	GAGCGGTGCTGTGGCAAGAGGGGACATAGCTCTCTGGGAGGGAAGAAATCCAAGAAG	738		
Db	121	GAGCGGTGCTGTGGCAAGAGGGGACATAGCTCTCTGGGAGGGAAGAAATCCAAGAAG	180		
Qy	739	AAGCGCAGCGAGTCAAGGGTCTCTCAGTGGCAGCAGCAAGAGAGGAAGAACTGGGG	798		
Db	181	AAGCGCAGCGAGTCAAGGGTCTCTCAGTGGCAGCAGCAAGAGAGGAAGAACTGGGG	240		
Qy	799	GGCTTGGGGAGGATGCCAACCGCAGGAGGAGGGGTGTGCAAGAGCATCGCCCTC	858		
Db	241	GGCTTGGGGAGGATGCCAACCGCAGGAGGAGGGGTGTGCAAGAGCATCGCCCTC	300		
Qy	859	CCACACAGCCCCCTGATGAGCTGTAGCCCGAGCTTAGTGTCTTGAATCAAGACCCCTG	918		
Db	301	CCACACAGCCCCCTGATGAGCTGTAGCCCGAGCTTAGTGTCTTGAATCAAGACCCCTG	360		
Qy	919	ACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCTCAGCAGGAGGAGGAGGAGCAG	978		
Db	361	ACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCTCAGCAGGAGGAGGAGGAGCAG	420		
Qy	979	GTCCAGCATCAGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1038		
Db	421	GTCCAGCATCAGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	480		
Qy	1039	AGGTGGAAGGATCTGGGGTGTCTGGGAGACGCGACCCCAAGAGGAGGAGGAGGAGCAG	1098		
Db	481	AGGTGGAAGGATCTGGGGTGTCTGGGAGACGCGACCCCAAGAGGAGGAGGAGGAGCAG	540		
Qy	1099	AAGCAGCTCTCTTTCTACACAGTCCCTCTCAGAGCTCCGGGTCTCCACCGAGCATCCCC	1158		
Db	541	AAGCAGCTCTCTTTCTACACAGTCCCTCTCAGAGCTCCGGGTCTCCACCGAGCATCCCC	600		
Qy	1159	AGGCTGAGATCCAGGCTCTGACATGGAGCTGAAGAGCATGAGGACATAGATGCTCA	1218		

Qy	1180	ACATGGAAGCTGAAGAGCATGAGGCACATAAGATGCTCACAGAGGCCCTTCAGCCAGG	1239
Db	301	ACATGGAAGCTGAAGAGCATGAGGCACATAAGATGCTCACAGAGGCCCTTCAGCCAGG	360
Qy	1240	AAGGACTCCGTGCAGCCTCAGCAGCCAGGCGCTGCCTCTTCCTTCCACCAAGCAATTCCTTT	1299
Db	361	AAGGACTCCGTGCAGCCTCAGCAGCCAGGCGCTGCCTCTTCCTTCCACCAAGCAATTCCTTT	420
Qy	1300	CTGCTGGTCTCTTGTTCGGATGGTAAATTCGAGAACTCTCCAGGACAAACTCGGCTGTGGCAC	1359
Db	421	CTGCTGGTCTCTTGTTCGGATGGTAAATTCGAGAACTCTCCAGGACAAACTCGGCTGTGGCAC	480
Qy	1360	AAAGGGGCTCGAGCGCCAGAGCCGAGCCAGCAGAGATCTGCAGAGAGGGCCACTGACCT	1419
Db	481	AAAGGGGCTCGAGCGCCAGAGCCAGAGCCAGCAGAGATCTGCAGAGAGGGCCACTGACCT	540
Qy	1420	AACCCGCCCTCGGAAGGCCAACTCGAGTTCCCGGTCCACCACCTCTCTCTGAGAGCGCCT	1479
Db	541	AACCCGCCCTCGGAAGGCCAACTCGAGTTCCCGGTCCACCACCTCTCTCTGAGAGCGCCT	600
Qy	1480	CATGCTCTGCCAGCCCTTCTCCAGGGCTACCAAGTAAACACCTTTTGGCCCTTCGGT	1539
Db	601	CATGCTCTGCCAGCCCTTCTCCAGGGCTACCAAGTAAACACCTTTTGGCCCTTCGGT	660
Qy	1540	TTGGTTCTCGGGTCCTCATCAGCCTCCA	1567
Db	661	TTGGTTCTCGGGTCCTCATCAGCCTCCA	688

Result No.	Score	Query #	Length	DB	ID	Description
1	50	2.7	1201	3	US-09-023-655-140	Sequence 140, App
2	36	2.0	1534	3	US-09-023-655-274	Sequence 274, App
C	33	1.8	1658	3	US-09-482-273-23	Sequence 23, Appl
	26	1.4	601	3	US-09-949-016-132624	Sequence 132624, A
	26	1.4	450395	3	US-09-949-016-15473	Sequence 15473, A
5	25	1.4	601	3	US-09-949-016-23380	Sequence 23380, A
6	25	1.4	601	3	US-09-949-016-141221	Sequence 141221, A
7	25	1.4	601	3	US-09-949-016-151342	Sequence 151342, A
8	25	1.4	601	3	US-09-949-016-151343	Sequence 151343, A
9	25	1.4	601	3	US-09-949-016-151405	Sequence 151405, A
10	25	1.4	601	3	US-09-949-016-151406	Sequence 151406, A
11	25	1.4	601	3	US-09-949-016-3399	Sequence 3399, Ap
C 12	25	1.4	1589	3	US-09-949-016-15158	Sequence 15158, A
C 13	25	1.4	9087	3	US-09-949-016-13251	Sequence 13251, A
14	25	1.4	11211	3	US-09-949-016-13251	Sequence 13251, A
C 15	25	1.4	19601	3	US-09-949-016-15629	Sequence 15629, A
C 16	25	1.4	33753	3	US-09-949-016-15741	Sequence 15741, A
C 17	25	1.4	33756	3	US-09-949-016-12006	Sequence 12006, A
C 18	25	1.4	102520	3	US-09-949-016-17367	Sequence 17367, A
C 19	25	1.4	102526	3	US-09-949-016-12448	Sequence 12448, A
C 20	25	1.4	118382	3	US-09-949-016-15996	Sequence 15996, A
C 21	25	1.4	118382	3	US-09-949-016-15997	Sequence 15997, A
C 22	24	1.3	321	2	US-08-171-385-22	Sequence 22, Appl
C 23	24	1.3	321	3	US-08-361-441B-22	Sequence 22, Appl
C 24	24	1.3	340	2	US-08-171-385-27	Sequence 27, Appl

REF ID: A66997

100

; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15473  
; LENGTH: 450395  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(450395)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15473

Query Match 1.4%; Score 26; DB 3; Length 450395;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1814 TAACGTGTAATAAAAAAAAAAAAAAAAAA 1839  
Db 418347 TAACGTGTAATAAAAAAAAAAAAAAAAAA 418372

## RESULT 6

US-09-949-016-23380  
; Sequence 23380, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23380  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-23380

Query Match 1.4%; Score 25; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAAAAAAAAAAAAAA 1839  
Db 222 AACTGTAAAAAAAAAAAAAAAAA 246

## RESULT 7

US-09-949-016-141221  
; Sequence 141221, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 141221  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-141221

Query Match 1.4%; Score 25; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 141221  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-141221

Query Match 1.4%; Score 25; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAAAAAAAAAAAAAA 1839  
Db 222 AACTGTAAAAAAAAAAAAAAAAA 246

## RESULT 8

US-09-949-016-151342  
; Sequence 151342, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 151342  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-151342

Query Match 1.4%; Score 25; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAAAAAAAAAAAAAA 1839  
Db 350 AACTGTAAAAAAAAAAAAAAAAA 374

## RESULT 9

US-09-949-016-151343  
; Sequence 151343, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 151343  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-151343

Query Match 1.4%; Score 25; DB 3; Length 601;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151343
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-151343

Query Match          1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 365 AACTGTAAAAA 389

RESULT 10
US-09-949-016-151405
; Sequence 151405, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151405
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-151405

Query Match          1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 350 AACTGTAAAAA 374

RESULT 11
US-09-949-016-151406
; Sequence 151406, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 151406
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-151406

Query Match          1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 365 AACTGTAAAAA 389

RESULT 12
US-09-949-016-3999/c
; Sequence 3999, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3999
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-3999

Query Match          1.4%; Score 25; DB 3; Length 1589;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 1433 AACTGTAAAAA 1409

RESULT 13
US-09-949-016-15158/c
; Sequence 15158, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15158
; LENGTH: 9087
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (1)...(9087)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15158

Query Match      1.4%; Score 25; DB 3; Length 9087;
Best Local Similarity 100.0%; Pred.No. 0.93;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 1186 AACTGTAAAAA 1162
```

```
RESULT 14
US-09-949-016-13251
; Sequence 13251, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13251
; LENGTH: 11211
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(11211)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13251
```

```
Query Match      1.4%; Score 25; DB 3; Length 11211;
Best Local Similarity 100.0%; Pred.No. 0.93;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 11122 AACTGTAAAAA 11146
```

```
RESULT 15
US-09-949-016-15629/c
; Sequence 15629, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15629
; LENGTH: 19601
```

```
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15629

Query Match      1.4%; Score 25; DB 3; Length 19601;
Best Local Similarity 100.0%; Pred.No. 0.89;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 18157 AACTGTAAAAA 18133

Search completed: December 30, 2005, 09:11:04
Job time : 357 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 04:23:41 ; Search time 1382 Seconds  
(without alignments)  
11003.897 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacgagggggaagcgc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 9793542 seqs, 4134689005 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA\_Main:  
1: /cgm2\_6/ptodata/1/pubpna/us07\_PUBCOMB.seq.\*  
2: /cgm2\_6/ptodata/1/pubpna/us08\_PUBCOMB.seq.\*  
3: /cgm2\_6/ptodata/1/pubpna/us09A\_PUBCOMB.seq.\*  
4: /cgm2\_6/ptodata/1/pubpna/us09B\_PUBCOMB.seq.\*  
5: /cgm2\_6/ptodata/1/pubpna/us10A\_PUBCOMB.seq.\*  
6: /cgm2\_6/ptodata/1/pubpna/us10B\_PUBCOMB.seq.\*  
7: /cgm2\_6/ptodata/1/pubpna/us10C\_PUBCOMB.seq.\*  
8: /cgm2\_6/ptodata/1/pubpna/us10D\_PUBCOMB.seq.\*  
9: /cgm2\_6/ptodata/1/pubpna/us10E\_PUBCOMB.seq.\*  
10: /cgm2\_6/ptodata/1/pubpna/us11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	5.4	100	7	US-10-430-201-4789 Sequence 4789, Ap
2	100	5.4	100	7	US-10-430-201-4790 Sequence 4790, Ap
3	50	2.7	444	7	US-10-424-599-33803 Sequence 33803, A
4	50	2.7	580	3	US-09-777-564-1232 Sequence 1232, Ap
5	50	2.7	580	5	US-10-015-219-1232 Sequence 1232, Ap
6	50	2.7	901	7	US-10-262-839-69 Sequence 69, Appl
7	50	2.7	1201	7	US-10-641-643-140 Sequence 140, App
8	50	2.7	1378	6	US-10-126-103-139 Sequence 139, App
9	50	2.7	1378	7	US-10-431-096-139 Sequence 139, App
10	50	2.7	1596	7	US-10-262-839-67 Sequence 67, Appl
11	50	2.7	1675	5	US-10-028-072-135 Sequence 135, App
12	50	2.7	1675	5	US-10-140-808-135 Sequence 135, App
13	50	2.7	1675	5	US-10-121-049-135 Sequence 135, App
14	50	2.7	1675	5	US-10-123-904-135 Sequence 135, App
15	50	2.7	1675	5	US-10-140-470-135 Sequence 135, App
16	50	2.7	1675	5	US-10-175-746-135 Sequence 135, App
17	50	2.7	1675	5	US-10-176-918-135 Sequence 135, App
18	50	2.7	1675	5	US-10-176-921-135 Sequence 135, App
19	50	2.7	1675	5	US-10-137-865-135 Sequence 135, App
20	50	2.7	1675	5	US-10-140-474-135 Sequence 135, App
21	50	2.7	1675	5	US-10-142-431-135 Sequence 135, App
22	50	2.7	1675	5	US-10-143-114-135 Sequence 135, App
23	50	2.7	1675	5	US-10-142-419-135 Sequence 135, App

24	50	2.7	1675	5	US-10-123-262-135	Sequence 135, App
25	50	2.7	1675	5	US-10-142-423-135	Sequence 135, App
26	50	2.7	1675	5	US-10-121-050-135	Sequence 135, App
27	50	2.7	1675	5	US-10-141-755-135	Sequence 135, App
28	50	2.7	1675	5	US-10-143-032-135	Sequence 135, App
29	50	2.7	1675	5	US-10-123-108-135	Sequence 135, App
30	50	2.7	1675	5	US-10-123-236-135	Sequence 135, App
31	50	2.7	1675	5	US-10-123-261-135	Sequence 135, App
32	50	2.7	1675	5	US-10-140-921-135	Sequence 135, App
33	50	2.7	1675	5	US-10-140-928-135	Sequence 135, App
34	50	2.7	1675	5	US-10-121-045-135	Sequence 135, App
35	50	2.7	1675	5	US-10-123-292-135	Sequence 135, App
36	50	2.7	1675	5	US-10-123-903-135	Sequence 135, App
37	50	2.7	1675	5	US-10-124-819-135	Sequence 135, App
38	50	2.7	1675	5	US-10-124-822-135	Sequence 135, App
39	50	2.7	1675	5	US-10-140-498-135	Sequence 135, App
40	50	2.7	1675	5	US-10-160-498-135	Sequence 135, App
41	50	2.7	1675	5	US-10-124-824-135	Sequence 135, App
42	50	2.7	1675	5	US-10-127-825A-135	Sequence 135, App
43	50	2.7	1675	5	US-10-127-829A-135	Sequence 135, App
44	50	2.7	1675	5	US-10-127-835A-135	Sequence 135, App
45	50	2.7	1675	5	US-10-127-839A-135	Sequence 135, App

ALIGNMENTS

RESULT 1  
US-10-430-201-4789  
; Sequence 4789, Application US/10430201  
; Publication No. US20040162679A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Linheng  
; TITLE OF INVENTION: Method for Predicting Gene Potential and Cell Commitment  
; FILE REFERENCE: 40716 (IP-010)  
; CURRENT APPLICATION NUMBER: US/10/430,201  
; PRIOR FILING DATE: 2003-05-05  
; PRIOR APPLICATION NUMBER: US 60/370,114  
; PRIOR FILING DATE: 2002-05-03  
; NUMBER OF SEQ ID NOS: 4879  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4789  
; LENGTH: 100  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; US-10-430-201-4789

Query Match	5.4%	Score 100;	DB 7;	Length 100;
Best Local Similarity	100.0%;	Pred. No. 2.9e-39;		
Matches 100;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	454	CACAACCTAGTCCACAAGGGGTCAAGGTGGTGGATATCCCTATGAGTGTGGAAC	513	
DB	1	CACAACCTAGTCCACAAGGGGTCAAGGTGGTGGATATCCCTATGAGTGTGGAAC	60	
QY	514	GAGACCTCAGCAGAGGTGGTGGTCAAGTCAAGAGCAGTGTG	553	
DB	61	GAGACCTCAGCAGAGGTGGTGGTCAAGTCAAGAGCAGTGTG	100	

RESULT 2  
US-10-430-201-4790  
; Sequence 4790, Application US/10430201  
; Publication No. US20040162679A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Linheng  
; TITLE OF INVENTION: Method for Predicting Gene Potential and Cell Commitment  
; FILE REFERENCE: 40716 (IP-010)  
; CURRENT APPLICATION NUMBER: US/10/430,201  
; CURRENT FILING DATE: 2003-05-05  
; PRIOR APPLICATION NUMBER: US 60/370,114  
; PRIOR FILING DATE: 2002-05-03  
; NUMBER OF SEQ ID NOS: 4879

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4790
; LENGTH: 100
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-430-201-4790

Query Match          5.4%; Score 100; DB 7; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.9e-39; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 0;

Qy 454 CACAACCTAGTCCACAAAGGGGTCAAGTGGTGATGCGATATCCCTATGAGCTGTGGAAAC 513
Db 1 CACAACCTAGTCCACAAAGGGGTCAAGTGGTGATGCGATATCCCTATGAGCTGTGGAAAC 60

Qy 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTG 553
Db 61 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTG 100

RESULT 3
US-10-424-599-33803
; Sequence 33803, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 33803
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_130528C.1
US-10-424-599-33803

Query Match          2.7%; Score 50; DB 7; Length 444;
Best Local Similarity 100.0%; Pred. No. 3.5e-14; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 586 GAGGACTGGTACAGGAACCAACCAGGAGGAGACCTGACTGAATTCCTCTG 635
Db 150 GAGGACTGGTACAGGAACCAACCAGGAGGAGACCTGACTGAATTCCTCTG 199

RESULT 4
US-09-777-564-1232
; Sequence 1232, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSEQ for Window Version 4.0
; SEQ ID NO 1232
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(580)
```

```
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-1232

Query Match          2.7%; Score 50; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 3.4e-14; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 586 GAGGACTGGTACAGGAACCAACCAGGAGGAGACCTGACTGAATTCCTCTG 635
Db 353 GAGGACTGGTACAGGAACCAACCAGGAGGAGACCTGACTGAATTCCTCTG 402

RESULT 5
US-10-015-219-1232
; Sequence 1232, Application US/10015219
; Publication No. US20030008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1232
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 451_522, 541, 548, 553, 576, 577, 578, 579
; OTHER INFORMATION: n = A,T,C or G
US-10-015-219-1232

Query Match          2.7%; Score 50; DB 5; Length 580;
Best Local Similarity 100.0%; Pred. No. 3.4e-14; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 586 GAGGACTGGTACAGGAACCAACCAGGAGGAGACCTGACTGAATTCCTCTG 635
Db 353 GAGGACTGGTACAGGAACCAACCAGGAGGAGACCTGACTGAATTCCTCTG 402

RESULT 6
US-10-282-839-69
; Sequence 69, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
```

```

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICANT NUMBER: <Unknown>
;   FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Zeller, Karen J.
;   REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (650) 855-0555
;   TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 140:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1201 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: THPINOB01
;   CLONE: 030137
; SEQUENCE DESCRIPTION: SEQ ID NO: 140 :
US-10-641-643-140

      Query Match          2.7%; Score 50; DB 7; Length 1201;
      Best Local Similarity 100.0%; Pred.No. 3.2e-14;
      Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY    586 GAGGACTGGTACAGGAACCAACCAGGAGGAGCCTGACTGAATTCCTCTG 635
      |||||||
DB    421 GAGGACTGGTACAGGAACCAACCAGGAGGAGCCTGACTGAATTCCTCTG 470

RESULT 8
US-10-126-103-139
; Sequence 139, Application US/10126103
; Publication No. US20030224486A1
; GENERAL INFORMATION:
;   APPLICANT: Bristol-Myers Squibb Company
;   TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-
;   FILE REFERENCE: D0108.np
; CURRENT APPLICATION NUMBER: US/10/126,103
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/284,962
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/286,645
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/346,986
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 139
; LENGTH: 1378
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-126-103-139

      Query Match          2.7%; Score 50; DB 6; Length 1378;
      Best Local Similarity 100.0%; Pred.No. 3.1e-14;
      Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY    586 GAGGACTGGTACAGGAACCAACCAGGAGGAGCCTGACTGAATTCCTCTG 635
      |||||||
DB    602 GAGGACTGGTACAGGAACCAACCAGGAGGAGCCTGACTGAATTCCTCTG 651

RESULT 9
US-10-431-096-139
; Sequence 139, Application US/10431096
; Publication No. US2004008696A1
; GENERAL INFORMATION:
;   APPLICANT: Bristol-Myers Squibb Company
;   TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-
;   PATHWAY
```

```

; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuroSeqdist version 0.1
; SEQ ID NO 67
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (388)..(549)
US-10-262-839-67

Query Match          2.7%; Score 50; DB 7; Length 1596;
Best Local Similarity 100.0%; Pred. No. 3.1e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 50; Conservative 0;

QY      586 GAGGACTGGTACAGGAACCACCCAGGAGGAGGAGCTGACCTGAATTCCTCTG 635
         |||||||
DB       795 GAGGACTGGTACAGGAACCACCCAGGAGGAGGAGCTGACCTGAATTCCTCTG 844

RESULT 11
US-10-028-072-135
; Sequence 135, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

US-10-123-904-133

Query Match 2.7%; Score 50; DB 5; Length 1675;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-14;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 586 GAGGACTGGTACAGGAACACCAGGAGGAGGACCTGACTGAATTCCTCTG 635  
 DB 868 GAGGACTGGTACAGGAACACCAGGAGGAGGACCTGACTGAATTCCTCTG 917

RESULT 15

US-10-140-470-135  
 ; Sequence 135, Application US/10140470  
 ; Publication No. US2003002331A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C160  
 ; CURRENT APPLICATION NUMBER: US/10/140,470  
 ; CURRENT FILING DATE: 2002-05-06  
 ; Prior Application removed - See Palm or File Wrapper  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 135  
 ; LENGTH: 1675  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-140-470-135

Query Match 2.7%; Score 50; DB 5; Length 1675;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-14;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 586 GAGGACTGGTACAGGAACACCAGGAGGAGGACCTGACTGAATTCCTCTG 635  
 DB 868 GAGGACTGGTACAGGAACACCAGGAGGAGGACCTGACTGAATTCCTCTG 917

Search completed: December 30, 2005, 09:34:20  
 Job time : 1383 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 04:41:12 ; Search time 299 Seconds  
(without alignments)  
3224.269 Million cell updates/sec

Title: US-09-696-686-47  
Perfect score: 1839  
Sequence: 1 ggcacgagggaagcgc.....taaaaaaaaaaaaaa 1839

Scoring table: Oligo NUC  
Gapop 60.0, Gapext 60.0

Searched: 4172979 seqs, 262114271 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Listing first 45 summaries

Database : Published Applications NA New:  
1: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	2.7	1675	6	US-10-131-826A-135
2	25	1.4	25	7	US-11-121-849-665859
3	25	1.4	25	7	US-11-121-849-665860
4	25	1.4	1612	6	US-10-750-185-3528
5	25	1.4	1833	6	US-10-750-185-38459
6	25	1.4	193789	7	US-11-112-908-55
7	24	1.3	25	7	US-11-121-849-665854
8	24	1.3	14911	7	US-11-112-908-63
9	24	1.3	154548	7	US-11-121-086-33
10	24	1.3	157230	7	US-11-112-908-64
11	24	1.3	158692	7	US-11-121-086-30
12	24	1.3	170508	7	US-11-112-908-62
13	24	1.3	172649	7	US-11-121-086-36
14	24	1.3	173115	7	US-11-112-908-65
15	24	1.3	193363	7	US-11-112-908-32
16	23	1.3	25	7	US-11-121-849-665857
17	23	1.3	946	6	US-10-750-185-37066
18	23	1.3	1120	6	US-10-750-185-29060
19	23	1.3	1382	6	US-10-750-185-39793
20	23	1.3	1445	6	US-10-750-185-29927
21	23	1.3	2352	6	US-10-909-125-835
22	23	1.3	10166	7	US-11-147-606-3
23	23	1.3	172147	7	US-11-112-908-22

c	24	23	1.3	188682	7	US-11-112-908-23	Sequence 23, Appl
c	25	23	1.3	246960	7	US-11-121-086-8	Sequence 8, Appl
c	26	22	1.2	201	6	US-10-995-561-18671	Sequence 18671, A
c	27	22	1.2	201	6	US-10-995-561-35450	Sequence 35450, A
c	28	22	1.2	201	6	US-10-995-561-35535	Sequence 35535, A
c	29	22	1.2	201	6	US-10-995-561-43529	Sequence 43529, A
c	30	22	1.2	201	6	US-10-995-561-46985	Sequence 46985, A
c	31	22	1.2	201	6	US-10-995-561-47049	Sequence 47049, A
c	32	22	1.2	201	6	US-10-995-561-50304	Sequence 50304, A
c	33	22	1.2	201	6	US-10-995-561-50700	Sequence 50700, A
c	34	22	1.2	201	6	US-10-995-561-50702	Sequence 50702, A
c	35	22	1.2	201	6	US-10-995-561-50871	Sequence 50871, A
c	36	22	1.2	201	6	US-10-995-561-53197	Sequence 53197, A
c	37	22	1.2	201	6	US-10-995-561-58387	Sequence 58387, A
c	38	22	1.2	201	6	US-10-995-561-80946	Sequence 80946, A
c	39	22	1.2	201	6	US-10-995-561-82910	Sequence 82910, A
c	40	22	1.2	576	7	US-11-156-909-2	Sequence 2, Appl
c	41	22	1.2	598	6	US-10-750-185-27071	Sequence 27071, A
c	42	22	1.2	600	6	US-10-750-185-21696	Sequence 21696, A
c	43	22	1.2	729	6	US-10-750-185-54079	Sequence 54079, A
c	44	22	1.2	776	7	US-11-156-909-24	Sequence 24, Appl
c	45	22	1.2	778	6	US-10-986-501-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-10-131-826A-135  
; Sequence 135, Application US/10131826A  
; Publication No. US20050245730A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C128  
; CURRENT APPLICATION NUMBER: US/10131,826A  
; CURRENT FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: 60/049911  
; PRIOR FILING DATE: 1997-06-18  
; PRIOR APPLICATION NUMBER: 60/056974  
; PRIOR FILING DATE: 1997-08-26  
; PRIOR APPLICATION NUMBER: 60/059113  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059115  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059117  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059122  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059184  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059352  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/059588

; PRIOR FILING DATE: 1997-09-19  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 135  
; LENGTH: 1675  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-131-826A-135

Query Match 2.7%; Score 50; DB 6; Length 1675;  
Best Local Similarity 100.0%; Pred. No. 8.3e-11;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 586 GAGGACTGGTACAGGAACCCAGGAGGAGACCTGACTGAATTCCTCTG 635  
Db 868 GAGGACTGGTACAGGAACCCAGGAGGAGACCTGACTGAATTCCTCTG 917

RESULT 2  
US-11-121-849-665859  
; Sequence 665859, Application US/11121849  
; Publication No. US20050272080A1  
; GENERAL INFORMATION:  
; APPLICANT: John Palma  
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S  
; FILE REFERENCE: 3684.1  
; CURRENT APPLICATION NUMBER: US/11/121,849  
; CURRENT FILING DATE: 2005-05-03  
; PRIOR APPLICATION NUMBER: 60/567,949  
; PRIOR FILING DATE: 2004-05-03  
; NUMBER OF SEQ ID NOS: 673904  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 665859  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-11-121-849-665859

Query Match 1.4%; Score 25; DB 7; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 595 TACAGGACCCAGGAGGAGACC 619  
Db 1 TACAGGACCCAGGAGGAGACC 25

RESULT 3  
US-11-121-849-665860  
; Sequence 665860, Application US/11121849  
; Publication No. US20050272080A1  
; GENERAL INFORMATION:  
; APPLICANT: John Palma  
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S  
; FILE REFERENCE: 3684.1  
; CURRENT APPLICATION NUMBER: US/11/121,849  
; CURRENT FILING DATE: 2005-05-03  
; PRIOR APPLICATION NUMBER: 60/567,949  
; PRIOR FILING DATE: 2004-05-03  
; NUMBER OF SEQ ID NOS: 673904  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 665860  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-11-121-849-665860

Query Match 1.4%; Score 25; DB 7; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 606 CCAGGAGGAGACCTGACTGAATTC 630  
Db 1 CCAGGAGGAGACCTGACTGAATTC 25

RESULT 4  
US-10-750-185-35528  
; Sequence 35528, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 35528  
; LENGTH: 1612  
; TYPE: DNA  
; ORGANISM: Bovine 19866881230530  
US-10-750-185-35528

Query Match 1.4%; Score 25; DB 6; Length 1612;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839  
Db 1242 AACTGTAAAAA 1266

RESULT 5  
US-10-750-185-38459/c  
; Sequence 38459, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 38459  
; LENGTH: 1833  
; TYPE: DNA  
; ORGANISM: Bovine 19866880774774  
US-10-750-185-38459

Query Match 1.4%; Score 25; DB 6; Length 1833;  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839  
Db 1789 AACTGTAAAAA 1765

```
RESULT 6
US-11-112-908-55/c
; Sequence 55, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 55
; LENGTH: 193789
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-55

Query Match      1.4%; Score 25; DB 7; Length 193789;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1815 AACTGTAAAAA... 1839
Db 5980 AACTGTAAAAA... 5956

RESULT 7
US-11-121-849-665854
; Sequence 665854, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 665854
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-665854

Query Match      1.3%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 CCCCTATGAGCTGTGGAACGAGAC 518
Db 1 CCCCTATGAGCTGTGGAACGAGAC 24

RESULT 8
US-11-112-908-63/c
; Sequence 63, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 63
; LENGTH: 149111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-63

Query Match      1.3%; Score 24; DB 7; Length 149111;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1816 ACTGTAAAAA... 1839
Db 136117 ACTGTAAAAA... 136094

RESULT 9
US-11-121-086-33
; Sequence 33, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: FOULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 33
; LENGTH: 154548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-33

Query Match      1.3%; Score 24; DB 7; Length 154548;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1816 ACTGTAAAAA... 1839
Db 71640 ACTGTAAAAA... 71663

RESULT 10
US-11-112-908-64/c
; Sequence 64, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 63
; LENGTH: 149111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-63
```

```
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 64
; LENGTH: 157230
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-64

Query Match      1.3%; Score 24; DB 7; Length 157230;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1816 ACTGTAAAAA 1839
Db 83463 ACTGTAAAAA 83440

RESULT 11
US-11-121-086-30
; Sequence 30, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 30
; LENGTH: 158692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-30

Query Match      1.3%; Score 24; DB 7; Length 158692;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1816 ACTGTAAAAA 1839
Db 40153 ACTGTAAAAA 40176

RESULT 12
US-11-112-908-62/c
; Sequence 82, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
```

```
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 62
; LENGTH: 170508
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-62

Query Match      1.3%; Score 24; DB 7; Length 170508;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1816 ACTGTAAAAA 1839
Db 67219 ACTGTAAAAA 67196

RESULT 13
US-11-121-086-36
; Sequence 36, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36
; LENGTH: 172649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-36

Query Match      1.3%; Score 24; DB 7; Length 172649;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1816 ACTGTAAAAA 1839
Db 126745 ACTGTAAAAA 126768

RESULT 14
US-11-112-908-65/c
; Sequence 65, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 65
; LENGTH: 173115
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-65
```

Query Match 1.3%; Score 24; DB 7; Length 173115;  
 Best Local Similarity 100.0%; Pred.No. 0.62;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1816 ACTGTAAAAA 1839  
 Db 18974 ACTGTAAAAA 18951

RESULT 15

US-11-112-908-32  
 ; Sequence 32, Application US/11112908  
 ; Publication No. US20050260659A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Davis, Lisa M.  
 ; TITLE OF INVENTION: Breast Cancer Biomarkers  
 ; FILE REFERENCE: 04-164-US  
 ; CURRENT APPLICATION NUMBER: US/11/112,908  
 ; CURRENT FILING DATE: 2005-04-22  
 ; PRIOR APPLICATION NUMBER: US 60/564,758  
 ; PRIOR FILING DATE: 2004-04-23  
 ; PRIOR APPLICATION NUMBER: US 60/575,978  
 ; PRIOR FILING DATE: 2004-06-01  
 ; PRIOR APPLICATION NUMBER: US 60/631,702  
 ; PRIOR FILING DATE: 2004-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/633,826  
 ; PRIOR FILING DATE: 2004-12-07  
 ; NUMBER OF SEQ ID NOS: 511  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 32  
 ; LENGTH: 193363  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-11-112-908-32

Query Match 1.3%; Score 24; DB 7; Length 193363;  
 Best Local Similarity 100.0%; Pred.No. 0.61;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1816 ACTGTAAAAA 1839  
 Db 84066 ACTGTAAAAA 84089

Search completed: December 30, 2005, 09:39:32  
 Job time : 302 secs

**THIS PAGE BLANK (USPTO)**